UNIVERSITÉ DE MONTRÉAL

DÉTERMINATION DE L'IMPACT DU PENTACHLOROPHÉNOL SUR LA DIVERSITÉ BACTÉRIENNE DANS LES SOLS À PARTIR DE LA TECHNIQUE DE L'ÉLECTROPHORÈSE SUR GEL EN GRADIENT DÉNATURANT

JONATHAN LALANDE

DÉPARTEMENT DE GÉNIE CHIMIQUE

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION

DU DIPLÔME DE PHILOSOPHIAE DOCTOR

(GÉNIE CHIMIQUE)

JUIN 2013

© Jonathan Lalande, 2013.

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée:

DÉTERMINATION DE L'IMPACT DU PENTACHLOROPHÉNOL SUR LA DIVERSITÉ BACTÉRIENNE DANS LES SOLS À PARTIR DE LA TECHNIQUE DE L'ÉLECTROPHORÈSE SUR GEL EN GRADIENT DÉNATURANT

présentée par : LALANDE Jonathan

en vue de l'obtention du diplôme de : Philosophiae Doctor

a été dûment acceptée par le jury d'examen constitué de :

M. JOLICOEUR Mario, Ph.D., président

Mme. DESCHÊNES Louise, Ph.D., membre et directrice de recherche

M. VILLEMUR Richard, Ph.D., membre et codirecteur de recherche

M. HENRY Olivier, Ph.D., membre

M. JUCK David, Ph.D., membre

DÉDICACE

« *N'espère rien de l'homme s'il travaille pour sa propre vie et non pour son éternité.* »

Antoine de Saint-Exupéry

REMERCIEMENTS

Six années pour en arriver là! Six années parsemées de joies et de déceptions, de hauts et de bas. Afin de conclure – enfin diront certains – comme il se doit cette aventure pleine de rebondissement qu'est la vie de thésard, je me dois de remercier certaines personnes ayant été impliquées directement ou indirectement dans mon projet.

Évidemment, je dois placer tout en haut de cette liste Louise. D'abord pour m'avoir accepté comme étudiant à la maîtrise, puis au doctorat. Puis pour m'avoir donné l'opportunité de m'impliquer dans l'enseignement de plusieurs cours. Puis pour sa grande disponibilité, à chaque fois que le besoin s'en faisait sentir. Mais surtout, pour m'avoir permis d'orienter mon projet dans la direction que je croyais être la bonne, pour avoir toujours placé mes intérêts en premier lorsque venait le temps de prendre des décisions et pour m'avoir toujours fait confiance, malgré les résultats qui tardaient à arriver. Ça été un plaisir de travailler avec toi!

Je ne pourrais passer sous silence la contribution de Richard, qui m'a toujours apporté très rapidement l'aide ou les conseils dont j'avais besoin, tout en me donnant toute la latitude et le temps dont j'avais besoin afin de développer mes idées. J'ai beaucoup apprécié tout le travail que tu as fourni, notamment en fin de projet, afin d'accroître la qualité des publications que nous avons réussi à écrire.

Merci aux partenaires du CIRAIG, à la Fondation de Polytechnique, au Fonds québécois de la recherche sur la nature et les technologies et au Conseil de recherches en sciences naturelles et en génie du Canada pour leur soutien financier tout au long de ce projet.

Je voudrais aussi remercier Étienne et Charles pour leur curiosité, sans quoi une grosse partie de ce projet n'aurait pas été possible. Un gros merci supplémentaire, à Étienne seul cette fois, pour avoir répondu à mes nombreuses questions!

Je m'en voudrais d'oublier Lucie, qui m'a apporté une aide très précieuse au laboratoire mais aussi ce qu'on pourrait appeler un support psychologique et moral lorsque les choses allaient moins bien, et pas seulement au labo…

Dans la catégorie support moral, je ne pourrais passer sous silence l'apport de PO, qui a entendu plus souvent qu'à son tour parler de communautés bactériennes, même si ça ne fait pas nécessairement (ou nécessairement pas…) partie de ses intérêts dans la vie. Les allers et retours de l'aréna seront dorénavant plus reposants.

Je tiens également à souligner l'apport de deux personnes bien spéciales à mes yeux, Nadia et Gladys, qui ont partagé ma vie (à des moments différents, bien entendu) tout au long de cette aventure. Je sais que je n'ai pas été toujours facile à supporter, surtout vers la fin du projet…

Un grand merci à, entre autres, Dave, Oli, Gump, Gab, Sophie, Isa, Rosie, Alyson et Anne Lautier pour tous ces moments partagés, ainsi qu'au CIRAIG en entier, un groupe composé de personnes fantastiques.

Et, finalement, le plus grand des mercis à deux individus que je place dans le club sélect des gens les plus importants dans ma vie, même s'ils en doutent peut-être parfois; mes parents! Vous m'avez donné les moyens de réussir, vous m'avez permis d'apprécier toutes ces années passées aux d'études, vous m'avez donné les moyens de réaliser d'autres projets tout en étant étudiant, vous avez été présents (sauf dans le temps des impôts) et vous m'avez toujours supporté dans mes choix. Juste ça…

Oh…j'allais oublier…un grand merci à Zorro pour son amour inconditionnel!

RÉSUMÉ

Les communautés microbiennes dans les sols sont parmi les plus diversifiées sur Terre. Ce faisant, les microorganismes du sol assurent en majeure partie plusieurs fonctions des sols tels que le cycle des nutriments et la détoxification des écosystèmes terrestres. En affectant cette diversité, la contamination des milieux naturels constitue un risque important pouvant réduire la capacité des écosystèmes à résister et récupérer aux différentes perturbations qu'ils doivent subir.

La diversité des écosystèmes naturels est donc une richesse à préserver. Effectivement, il a été montré que les écosystèmes plus diversifiés étaient plus résistants et résilients aux perturbations naturelles et anthropiques. Depuis le début de l'ère industrielle, la diversité des écosystèmes naturels est en constant déclin à cause, notamment, de la contamination des sols, de l'air et des cours d'eau. Afin de prédire l'effet d'une substance sur une communauté biologique et ainsi contrôler ou limiter son utilisation, il est nécessaire de produire de l'information toxicologique sur une large gamme d'organismes. Les données écotoxiques représentatives des différentes espèces testées seront colligées et utilisées afin d'estimer l'effet d'un contaminant sur toute une communauté. Lorsque le centre d'intérêt est une communauté biologique évoluant en milieu terrestre, force est de constater que la disponibilité de ces indicateurs est très problématique. Les communautés microbiennes des sols sont nécessaires au bon fonctionnement, à long terme, des écosystèmes terrestres. Or, ces dernières sont virtuellement absentes des bases de données écotoxiques. L'absence de données pour ce groupe très important n'est pas étrangère au fait que l'étude d'organismes aussi petits pose des difficultés techniques particulières. Ainsi, la majorité des publications s'étant intéressées à l'effet des contaminants sur les microorganismes du sol ont utilisé des indicateurs fonctionnels liés directement ou indirectement à ces organismes. Les indicateurs les plus utilisés sont les mesures de biomasse microbienne, d'activité microbienne et d'activité enzymatique. Bien que très utiles afin de déterminer l'état de santé d'un sol, ces indicateurs ont montré certaines limites importantes. Puisque les mesures fonctionnelles sont très intégratrices (toutes les espèces y participent), plusieurs auteurs ont noté que certains de ces indicateurs avaient le potentiel d'augmenter suite à une contamination. En fractionnant les communautés microbiennes en plusieurs sous-parties (les espèces), les indices de diversité, qui considèrent à la fois le nombre d'espèces présentes et l'abondance relative de chaque espèce, pourraient être beaucoup plus sensibles à une contamination que les indicateurs fonctionnels. Par contre, il est techniquement difficile d'estimer précisément la diversité des communautés

microbiennes des sols. D'abord, les méthodes moléculaires utilisées afin d'étudier la diversité microbienne des sols ne permet pas de cibler simultanément tous les microorganismes du sol. Des trois sous-groupes composant les communautés microbiennes des sols (bactéries, archées et champignons), les bactéries sont les plus diversifiées et sont le plus souvent ciblées. Ensuite, il a été montré que les populaires techniques de profilage, qui ne donnent de l'information que sur les espèces les plus abondantes, sont incapables de produire des estimés de diversité corrélés à la diversité bactérienne réelle des sols. Les nouvelles technologies de séquençage d'ADN récemment arrivées sur le marché ont le potentiel de permettre l'étude quantitative de la diversité bactérienne des sols. Par contre, plusieurs auteurs ont montré que certaines difficultés techniques pourraient générer des distorsions dans les estimés de diversité produits par ces plateformes, bien que l'importance quantitative de ces distorsions n'ait jamais été clairement établie. Aussi, la disponibilité de ces appareils peut être problématique pour plusieurs chercheurs.

Plusieurs limitations ont été identifiées afin d'estimer l'effet des contaminants dans les écosystèmes terrestres. L'information toxicologique pour les espèces terrestres est très limitée et ne concerne souvent que quelques espèces. Considérant qu'ils sont virtuellement absents des bases de données écotoxiques, la génération de données de toxicité spécifiques aux bactéries du sol constitue un moyen permettant de s'attaquer à cette problématique. Vu l'importance de ces communautés dans les écosystèmes, tant du point de vue de leur diversité que de celui de leur importance fonctionnelle, ces données constitueraient de l'information hautement pertinente afin de mieux comprendre l'effet des contaminants en milieux terrestres. Pour ce faire, une mesure de toxicité basée sur la diversité génétique des communautés bactériennes dans les sols présente plusieurs avantages par rapport aux indicateurs fonctionnels souvent utilisés par le passé. Par contre, l'utilisation de la diversité microbienne en tant qu'indicateur de toxicité présente également des difficultés, la plus importante étant la capacité des techniques actuelles à produire des estimés quantitatifs de diversité. Il a été évoqué qu'une technique de profilage très répandue, l'électrophorèse sur gel en gradient dénaturant (DGGE), pourrait posséder cette capacité. Donnant généralement de l'information uniquement sur les espèces les plus abondantes des communautés étudiées, il a été montré que cette technique pourrait également contenir de l'information sur leur richesse. Par contre, aucune approche n'a pour le moment été développée afin d'utiliser cette information. De plus, l'analyse des profils DGGE présente des difficultés particulières lorsque des

communautés très diversifiées sont étudiées. Dès lors, il est difficile d'affirmer que l'information produite sur les espèces les plus abondantes des communautés est quantitativement représentative.

Afin de développer une approche permettant d'utiliser toute l'information extraite des profils DGGE, le pentachlorophénol semble être une substance idéale. En effet, ce contaminant se retrouve fréquemment dans les écosystèmes naturels canadiens, possède des propriétés physico-chimiques susceptibles de faire varier sa toxicité en fonction de la composition du sol utilisé et a été testé sur plusieurs espèces terrestres différentes. De plus, son effet sur les microorganismes du sol a fait l'objet de quelques publications. Il sera intéressant de comparer la réponse et la sensibilité d'un indicateur basé sur la diversité bactérienne des sols à des mesures biochimiques. À cette fin, la mesure du potentiel d'activité hydrolytique de la fluorescéine diacétate, un essai simple produisant des résultats corrélés aux mesures de biomasse microbienne, semble être idéale. Effectivement, la biomasse microbienne s'est avérée être un indicateur sensible au pentachlorophénol dans quelques études différentes.

Sur la base des limitations identifiées, les objectifs principaux de ce projet sont les suivants :

1. Proposer une approche analytique permettant d'extraire des profils DGGE de l'information quantitative sur la dominance et la richesse des communautés étudiées;
2. Développer, à partir de profils DGGE *in silico*, un modèle empirique permettant de produire des estimés de diversité corrélés à la diversité des communautés étudiées à partir de l'information extraite des profils DGGE;
3. Valider ce modèle sur des échantillons réels en comparant la diversité estimée à partir du DGGE à celle produite par une nouvelle technologie de séquençage d'ADN;
4. Quantifier l'effet du pentachlorophénol sur la diversité bactérienne des sols et comparer cette réponse à celle basée sur la mesure du potentiel d'activité hydrolytique de la fluorescéine diacétate des sols.

Sur la base de ces objectifs, une approche analytique a été développée à partir de profils DGGE *in silico* à l'aide du logiciel Matlab. Contrairement aux logiciels d'analyse de gels commerciaux, qui soustraient le bruit de fond à l'aide d'une bille virtuelle roulant sous les profils DGGE, cette approche utilise un profil de bruit de fond défini manuellement par l'utilisateur. Le bruit de fond ainsi soustrait est utilisé afin de définir un nouveau paramètre, le ratio pic-sur-signal, qui contient

l'information sur la richesse des communautés étudiées. Une fois le bruit de fond soustrait, l'information sur les espèces les plus abondantes de la communauté est générée à l'aide d'un algorithme d'optimisation facilitant l'analyse de profils complexes.

À partir des informations extraites des gels DGGE (abondance relative des espèces les plus abondantes et richesse des communautés), un modèle empirique a été développé afin de produire des estimés de diversité quantitatifs à partir du DGGE. La distribution rang-abondance produite par l'analyse des bandes DGGE est premièrement multipliée par le ratio pic-sur signal extrait du profil. Par la suite, en utilisant un modèle d'abondance empirique, la distribution loi de puissance, un modèle d'élongation des distributions rang-abondance DGGE a été développé. Ce modèle a été paramétré en observant les distributions rang-abondance des communautés bactériennes *in silico* utilisées afin de construire les profils analysés. Puisque le paramétrage du modèle dépend du ratio pic-sur-signal de l'échantillon analysé, le processus d'élongation est propre à chaque échantillon.

Afin de valider ce modèle d'élongation, les communautés bactériennes des mêmes échantillons de sol ont été étudiées à partir du DGGE et de l'Ion Torrent, une nouvelle technologie de séquençage. Deux sols de texture différente ont été utilisés afin de valider ce modèle. Chaque sol a été contaminé avec du pentachlorophénol, à trois concentrations différentes (en plus de l'échantillon non contaminé).

Afin d'étudier l'effet écotoxique du pentachlorophénol sur la diversité bactérienne des sols, les deux sols mentionnés ci-haut ont été contaminés cette fois à sept concentrations différentes, en triplicata. En parallèle, l'activité hydrolytique de la fluorescéine diacétate a été mesurée dans tous les échantillons de sol. La diversité et l'activité hydrolytique de tous les échantillons ont été utilisées afin de produire des courbes concentration-réponse représentatives de l'effet du PCP sur les communautés bactériennes (diversité) ou, plus généralement, microbienne (activité de la fluorescéine diacétate). Aussi, en utilisant les données écotoxiques terrestres disponibles dans les bases de données pour le pentachlorophénol, une courbe de sensibilité des espèces a été tracée pour ce contaminant. Les indicateurs écotoxiques extraits des courbes concentration-réponse et de la courbe de sensibilité des espèces ont été comparés.

En travaillant sur des profils DGGE *in silico*, il a été montré que la représentativité des résultats générés par les approches analytiques généralement employées pour analyser les gels DGGE et implémentées dans trois logiciels commerciaux variait d'un échantillon à l'autre. Au contraire,

l'approche analytique développée dans le cadre de ce projet a permis d'extraire des profils DGGE des ratios pics-sur signal précis et de produire des patrons de dominance représentatifs des communautés étudiées. Il a été montré que ces patrons de dominance étaient très semblables à ceux générés par le groupement des séquences utilisées pour construire les profils DGGE si un pourcentage de similitude de 98% était utilisé.

Le modèle d'élongation calibré à partir des distributions rang-abondance produites en groupant les séquences d'ADN avec un pourcentage de similitude de 98% a montré un très fort potentiel afin de permettre l'estimation de la diversité bactérienne à partir du DGGE. Ainsi, il a été possible d'estimer la diversité réelle des communautés étudiées avec une précision généralement supérieure à ± 5%. Une corrélation très significative existait entre les indices de diversité réels et ceux estimés à partir du DGGE ($R^2 > 0,99$).

En utilisant deux approches différentes afin d'étudier les mêmes communautés bactériennes, il a été possible d'identifier des distorsions dans les résultats de chacune des approches. Ainsi, il a été montré que le fait d'injecter une grande quantité d'ADN (~ 400 ng) dans les puits DGGE risquait de causer une saturation locale du gel. Cette saturation, qui survient dans les régions du gel où une grande quantité d'ADN migre, peut empêcher les bandes DGGE les plus brillantes d'atteindre leurs intensités maximales. Injecter moins d'ADN (~ 200 ng) a permis de régler les problèmes liés à la saturation des gels. La situation s'est avérée être plus complexe pour les résultats générés par l'Ion Torrent. Utilisée à la limite de ses capacités, la qualité des résultats de séquençage produits par cette plateforme fut problématique. Ainsi, des distorsions liées à la faible profondeur de séquençage atteinte et aux erreurs de séquençage furent identifiés. De plus, il a été qualitativement montré que le pourcentage GC des séquences risquait de causer des distorsions intra-échantillons dans l'abondance relative des espèces. Le fait de réduire l'influence de ces distorsions sur les estimés de diversité produits par chaque approche a permis l'obtention de corrélations significatives ($R^2 > 0,81$) entre les deux séries de données. Lorsqu'utilisées à des fins comparatives, ces deux technologies peuvent donc mener à des résultats quantitativement similaires.

Lors de la construction des courbes concentration-réponse à partir des indices de diversité mesurés pour les différents échantillons de sol, il a été possible de remarquer que les échantillons de sol les plus contaminés n'étaient pas les moins diversifiés. Cette observation peut s'expliquer de deux façons différentes. Le pentachlorophénol a pu avoir un effet toxique sur toutes les espèces

bactériennes, empêchant ainsi l'apparition de bandes DGGE brillantes sur les profils et permettant par le fait même à certaines espèces d'être suffisamment abondante (proportionnellement) afin de produire une bande visible sur le gel. Alternativement, ou parallèlement, il est possible que le temps d'incubation de 28 jours n'ait pas permis la dégradation de l'ADN des bactéries tuées par l'ajout de pentachlorophénol. Malgré ces observations, les indices de diversité se sont avérés être beaucoup plus sensibles (plus de 25 fois) au pentachlorophénol que les mesures d'activité hydrolytique de la fluorescéine diacétate. De plus, les courbes concentration-réponse tracées à l'aide des mesures de diversité se sont avérées très similaires à la courbe de sensibilité des espèces représentative du pentachlorophénol.

Ainsi, cette thèse a permis le développement d'une méthodologie permettant de comparer la diversité de communautés bactériennes de différents sols à l'aide du DGGE, une approche très répandue qui n'avait préalablement pas cette capacité. De plus, il a été possible de montrer que les indices de diversité sont plus sensibles au pentachlorophénol que certains indicateurs fonctionnels, et que la réponse des communautés bactériennes face à une exposition au pentachlorophénol était très similaire à la réponse des communautés macroscopiques terrestres exposées à la même substance. Si ces observations peuvent être généralisées sur plus de contaminants, l'approche proposée dans cette thèse constituera un outil excessivement puissant afin de mieux comprendre l'effet des contaminants ou des mélanges complexes sur les communautés biotiques terrestres. Et si cette similitude est tout simplement fortuite, le cadre de travail présenté dans cette thèse sera tout de même en mesure de produire de l'information écotoxicologique sur les communautés bactériennes, archéennes et fongiques des sols, trois groupes absents des bases de données écotoxicologiques.

ABSTRACT

Soil microbial communities are among the most diversified on Earth and play key roles in many soil-mediated functions such as biogeochemical cycles and xenobiotics degradation. Affecting this diversity, soil contamination constitutes an important risk having the potential to reduce the capacity of terrestrial ecosystems to resist and recover from the many natural and anthropogenic perturbations that they must undergo.

As many authors have shown, diversified ecosystems are more resistant and resilient to perturbations and diversity is an asset that must be preserved. However, since the beginning of the industrial era, soil, air and water contamination leads to a constant decrease in natural ecosystems' diversity. In order to protect ecosystems from diversity losses caused by xenobiotics, ecotoxicological information about many different species is required. This information on single species will be compiled and used to draw a species sensitivity distribution that allows estimating the effect of a contaminant on a whole ecosystem. For terrestrial ecosystems, data scarcity is a major limitation to the application of this framework. Although necessary to the long term functioning of terrestrial ecosystems, soil microbial communities are virtually absent from terrestrial ecotoxicological databases. A major reason for that is linked with the technical difficulties in studying such small organisms. Most of the publications that focused on the effect of xenobiotics to microbial communities measured functional parameters directly or indirectly linked to soil microorganisms, such as microbial biomass or activity and enzymes activity. Although highly informative of soil health, these indicators showed important limitations when used as toxicity indicators. Being very integrative, those measures often showed increases after soil contamination, especially if the tested contaminant is biodegradable. Fractionating microbial communities into smaller distinct parts – the species – ecological diversity indices that consider simultaneously the richness of the community (number of species) and the relative abundance of the different species could be more sensitive to xenobiotics than functional parameters. However, it is still technically difficult to accurately estimate the diversity of soil microbial communities. First of all, the molecular approaches used to study microbial diversity are not able to consider at once all the microorganisms (bacteria, archaea and fungi) composing these communities. As a consequence, bacterial communities – the most diversified of these subgroups – are selected by most of the authors. Then, it was shown that the widespread profiling methods – only able to yield quantitative information about the relative abundance of the most abundant species – are unable to

produce diversity estimates correlated to the actual diversity of soil bacterial communities. The newer next-generation DNA sequencing technologies theoretically allow studying the diversity of bacterial communities with an unprecedented depth. However, many authors showed that diversity estimates produced from sequencing results may be biased to an unknown extent. Furthermore, sequencer availability may still be problematic for many researchers.

Many limitations hampering the estimation of the effect of xenobiotics in terrestrial ecosystems have been identified. Terrestrial ecotoxicological data are scarce and do not concern enough different species. Being almost absent in the ecotoxicological databases, generating data about soil bacterial communities allows us to tackle this problem. Furthermore, considering their importance in terrestrial ecosystems, generating information about the effect of xenobiotics to these communities would be of high environmental relevance. To do so, ecological diversity indices present many advantages compared with functional indicators, but also many difficulties. Being able to accurately estimate the diversity of these very complex communities is, without a doubt, the major one. In one publication, it was shown that denaturing gradient gel electrophoresis (DGGE) – a widespread profiling method – is theoretically able to yield quantitative information on the relative abundance of the dominant species, but also on the richness of the studied communities. However, no framework has been developed in order to use this information. Furthermore, the analysis of complex DGGE migration profiles of soil bacterial communities presents many difficulties. It is therefore hard to say that the information extracted from the DGGE migration profiles can be used to quantitatively study bacterial community diversity.

In order to develop a framework allowing quantifying the effect of xenobiotics on the diversity of soil bacterial communities, pentachlorophenol seems to be an ideal case study. This contaminant is ubiquitous in Canadian terrestrial ecosystems, possesses interesting physico-chemical properties potentially allowing this substance to exhibit different toxicities in different soil types and, most of all, was tested on many different terrestrial species. Furthermore, its effect on soil microorganisms was already studied by some authors. It will be interesting to compare the sensitivity of ecological diversity indices and functional measures to detect pentachlorophenol. To do so, the assay of fluorescein diacetate hydrolytic activity seems to be perfectly suited; it is a simple assay producing results well correlated with microbial biomass measures that were shown by some authors to be sensitive to pentachlorophenol contamination.

On the basis of the above-mentioned limitations, the projects' main objectives are the following:

1. Develop an analytical approach allowing the extraction from DGGE migration profiles information on the relative abundance of the dominant species and on the richness of the studied communities;
2. Using *in silico* DGGE migration profiles, develop an empirical framework allowing the accurate estimation of the diversity of soil bacterial communities using the information extracted from the profiles;
3. Validate the framework on real samples by comparing the diversity estimates produced by DGGE and by a next-generation sequencing technology;
4. Quantify and compare the effect of pentachlorophenol on the diversity of soil bacterial communities and on the soils' fluorescein diacetate hydrolytic activity.

Based on these objectives, an analytical approach was developed on *in silico* DGGE profiles using the software Matlab. Unlike commercial software programs that use a virtual rolling disk to subtract the DGGE profile background noise, this framework uses a background profile manually adjusted by the analyst. This background noise is used to define a new parameter, the peak-to-signal ratio that was proved to be related to community richness. Then, the information about the relative abundance of the dominant species is extracted using a semi-automated algorithm, thus facilitating the analysis of complex DGGE profiles.

Using the information extracted from the profiles, an empirical framework allowing the accurate estimation of the diversity of soil bacterial communities was developed. The partial rank-abundance distributions produced by the DGGE bands quantification step are first of all normalized using the peak-to-signal ratios extracted from the profiles. Then, using the power law abundance model, an elongation model was parameterized using the complete rank-abundance distributions used to synthesize the *in silico* DGGE profiles. Because the model parameters depend on the peak-to-signal ratios, the elongation process is therefore sample specific.

To further validate this elongation model, the diversity of the bacterial communities of eight different soil samples were studied using both DGGE profiling and Ion Torrent sequencing, a next-generation sequencing platform. Two differently textured soils were used, a loamy and a sandy

soil. For each soil, four subsamples were prepared; one was kept clean and the three others were contaminated with different concentrations of pentachlorophenol.

The ecotoxicological effects of pentachlorophenol on the diversity of soil bacterial communities was further studied using the two same soils. This time, seven different pentachlorophenol concentrations were used and all the samples were prepared in triplicate. The samples' fluorescein diacetate hydrolytic activity was also measured. Concentration-response modeling was conducted on the two types of indicators (diversity and activity). Furthermore, ecotoxicological databases were screened in order to find data on the toxicity of pentachlorophenol to terrestrial species. These data were used to draw a species sensitivity distribution – representative of the range of sensitivity of terrestrial species exposed to pentachlorophenol – that was compared with the above-mentioned concentration-response curves.

Working on *in silico* DGGE profiles, it was shown that the representativeness of the results produced by commercial gel analysis software programs was sample-dependent. On the contrary, the Matlab-based framework presented in this thesis allowed the extraction of accurate peak-to-signal ratio values and the production of representative dominance profiles for all the analyzed samples. It was shown that these dominance profiles were very similar to those produced by DNA sequences clustering algorithms when using a similarity level of 98%.

The elongation model calibrated with the sequencing datasets used to synthesize the *in silico* DGGE profiles showed great potential. Calibrated in order to reproduce the rank-abundance distributions generated by sequencing datasets clustering at 98% similarity, the framework allowed the estimation of the diversity of the studied communities with a precision generally higher than ± 5%. A very significant correlation was found between those DGGE-based diversity estimates and true diversity at a similarity level of 98% ($R^2 > 0.99$).

Using two different techniques to study the same bacterial communities allowed identifying biases in the results of both technologies. For DGGE, it was clearly shown that injecting too much DNA in the wells (~ 400 ng) could cause localized gel saturation, hampering the brightest peaks to reach their full intensity and thus underestimating the dominance of the communities. Injecting less DNA (~ 200 ng) was a simple remedy against this problem. The situation was much more complex for Ion Torrent sequencing. Used at the limit of its capacity, the quality of the dataset produced by this platform was low. As a consequence, biases linked with sampling depth and sequencing errors

were identified. Furthermore, it was qualitatively shown that GC content considerations were susceptible to cause intra-sample distortions in the relative abundance of species. Furthermore, all these biases were sample-dependent. Reducing the quantitative importance of some biases allowed finding a significant correlation between the diversity estimates produced by DGGE and Ion Torrent sequencing ($R^2 > 0.81$). For comparative diversity studies, these two approaches can therefore yield similar ecological conclusions.

While constructing concentration-response curves from the ecological diversity indices, it was observed that the most contaminated soils were not the least diversified ones. Such a result could happen if the toxic effect of pentachlorophenol was so high that even the tolerant species were unable to grow, thus hampering the apparition of bright bands on the DGGE profile. At the same time, it is also possible that pentachlorophenol induced a high bacterial mortality and that the DNA of these organisms, released in the soil, was not degraded during the 28 day incubation period. Notwithstanding the above, ecological diversity indices proved to be at least 25 times more sensitive to pentachlorophenol than the soils' fluorescein diacetate hydrolysis activity, meaning that structural changes could happen before functional losses are observed. The most promising result was, without a doubt, a very good correspondence that was found between the diversity-based concentration-response curves and the species sensitivity distribution that was observed for pentachlorophenol.

This thesis allowed developing a complete framework allowing the comparison of soil bacterial community diversity using DGGE, a widespread technique that was unable to do so before. Furthermore, it was shown that diversity-based indicators were more sensitive to pentachlorophenol than some integrative functional indicators and that the response of soil bacterial communities to pentachlorophenol was highly similar to the response of macroscopic terrestrial communities exposed to the same substance. If this similarity can be established for other contaminants or classes of contaminants, this methodology has the potential to help fill many gaps in terrestrial ecotoxicological databases and in the understanding of the effect of xenobiotics or complex mixtures on terrestrial ecosystems. And if these similarities are fortuitous, the framework can still provide relevant information about the toxicity of xenobiotics to organisms poorly represented in the ecotoxicological databases; soil bacterial, archeal and fungal communities.

TABLE DES MATIÈRES

DÉDICACE ..... *III*

REMERCIEMENTS ..... *IV*

RÉSUMÉ ..... *VI*

ABSTRACT ..... *XII*

TABLE DES MATIÈRES ..... *XVII*

LISTE DES TABLEAUX ..... *XX*

LISTE DES FIGURES ..... *XXII*

LISTE DES SIGLES ET ABRÉVIATIONS ..... *XXV*

INTRODUCTION *1*

CHAPITRE 1 REVUE DE LITTÉRATURE *3*

1.1 Diversité biologique dans les écosystèmes naturels **3**

1.1.1 État et tendances 3

1.1.2 Lien biodiversité – fonction – stabilité 5

1.1.3 Quantification de la biodiversité 5

1.2 Évaluation de l'effet des contaminants sur les populations et les écosystèmes **9**

1.2.1 Essais de toxicité sur les populations 9

1.2.2 Évaluation de la sensibilité des écosystèmes 14

1.2.3 Particularités des écosystèmes terrestres 16

1.3 Importance, structure et diversité des communautés microbiennes des sols **19**

1.4 Évaluation de l'effet des contaminants sur la communauté microbienne d'un sol **20**

1.4.1 Approches fonctionnelles 21

1.4.2 Approches moléculaires 22

1.5 Le pentachlorophénol comme cadre d'étude **31**

CHAPITRE 2 PRÉSENTATION GÉNÉRALE DES TRAVAUX *33*

2.1 Problématique **33**

2.2 Hypothèse de recherche et objectifs 34

2.3 Méthodologie générale 35

2.4 Présentation des manuscrits 36

CHAPITRE 3 DÉVELOPPEMENT D'UNE APPROCHE ANALYTIQUE PERMETTANT L'EXTRACTION D'INFORMATIONS QUANTITATIVES D'UN PROFIL DGGE 38

3.1 Délimitation et soustraction du bruit de fond 38

3.2 Quantification des pics 41

CHAPITRE 4 DÉVELOPPEMENT D'UNE MÉTHODOLOGIE PERMETTANT DE QUANTIFIER LA DIVERSITÉ DES COMMUNAUTÉS BACTÉRIENNES DES SOLS À PARTIR DU DGGE 44

4.1 Présentation du premier manuscrit 44

4.2 "A new framework to accurately quantify soil bacterial community diversity from DGGE" . 45

4.2.1 Abstract 45

4.2.2 Introduction 45

4.2.3 Methods 48

4.2.4 Results 55

4.2.5 Discussion 62

4.2.6 Acknowledgements 67

CHAPITRE 5 VALIDATION DE LA MÉTHODOLOGIE D'ÉLONGATION DES RAD DGGE SUR DES ÉCHANTILLONS DE SOL CONTAMINÉS AU PENTACHLOROPHÉNOL 68

5.1 Lien entre le premier et le second manuscrit 68

5.2 "Can DGGE and Ion Torrent sequencing yield similar quantitative conclusions when comparing the diversity of soil bacterial communities?" 69

5.2.1 Abstract 69

5.2.2 Introduction 69

5.2.3 Methods 71

5.2.4 Results 77

5.2.5 Discussion 88

5.2.6 Acknowledgements 93

CHAPITRE 6 DÉTERMINATION DE L'EFFET ÉCOTOXIQUE DU PENTACHLOROPHÉNOL SUR LA DIVERSITÉ DES COMMUNAUTÉS BACTÉRIENNES DES SOLS 94

6.1 Lien entre le deuxième et le troisième manuscrit 94

6.2 "Assessing the impact of pentachlorophenol on soil bacterial communities using ecological diversity indices" 95

6.2.1 Abstract 95

6.2.2 Introduction 95

6.2.3 Material and methods 97

6.2.4 Results 103

6.2.5 Discussion 109

6.2.6 Acknowledgements 112

CHAPITRE 7 DISCUSSION GÉNÉRALE 113

CONCLUSION 119

Contributions scientifiques 119

Contributions pratiques 120

Perspectives et recommandations 120

BIBLIOGRAPHIE 123

ANNEXES 144

LISTE DES TABLEAUX

Tableau 1-1 : Nombre d'espèces décrit et estimé pour les trois domaines biologiques. Tiré de Groombridge et Jenkins (2002). 4

Tableau 1-2 : Liste non-exhaustive des indices de diversité utilisés afin de quantifier la diversité des communautés biotiques. S correspond au nombre d'espèces observées, p_i correspond à l'abondance relative de l'espèce de rang i et N correspond au nombre d'individus contenus dans l'échantillon. 7

Tableau 1-3 : Différences entre type d'indicateur et fonction suivie pour différents organismes exposés au cuivre. Adapté de Roman et al. (2007). Toutes les valeurs sont en mg de cuivre/kg de sol. 13

Tableau 1-4 : Nombre de contaminants et types d'indicateurs couverts par la base de données ECOTOX (U.S. Environmental Protection Agency, 2013) pour les écosystèmes aquatiques et terrestres. Si la base de données contenait plus d'une entrée pour une même combinaison contaminant – organisme - type d'indicateur (i.e. une NOEC pour la reproduction et une NOEC pour la mortalité), cette combinaison a été comptabilisée une seule fois. 18

Tableau 4-1 : For the Matlab-based framework, TotalLab Quant, GelCompar II and BIO-1D, optimal parameters obtained for all the samples. Ball size and similarity level values were selected as those simultaneously generating ΔPSR, ΔH' and Δ1/D values lower than ±10% over the widest range of cut-off values. 57

Tableau 4-2 : Parameters that made it possible to minimize indicator values when simultaneously considering all the samples. 59

Tableau 4-3 : Deviation of DGGE-based diversity estimates from the indices calculated using the untruncated true RADs. DGGE-based indices were calculated solely from the quantified peaks (with a cut-off of 1.0%) or using the elongated RADs. Results are presented as the average, minimum and maximum of all nine samples, as well as the slope and coefficient of determination generated by linear regression analysis. 62

Tableau 5-1 : Physico-chemical properties of the soils used in the study. 71

Tableau 5-2 : For all the samples, number of reads matching both primers before and after processing. The deletion percentage corresponding to the eight samples are also presented. 81

Tableau 5-3 : Potential influence of sequencing errors on the indices calculated from sequencing datasets. The values correspond to the relative index variation (in %) when the pseudo-single linkage clustering algorithm (Huse et al., 2010) is applied using a dissimilarity of 2 or 4 bases instead of 3. 84

Tableau 5-4 : Slopes and coefficients of correlation of the linear regressions conducted between Ion Torrent sequencing and DGGE-based diversity estimates. All the coefficients were determined by forcing the intercept of the regression line to be the origin. For the Ion Torrent data, uncorrected refers to the diversity indices

calculated directly from the clustered datasets, normalized corresponds to the values produced by sampling all the datasets 5,000 times and corrected values were produced with the DGGE RADs sampling framework. The numbers 2, 3 and 4 refer to the dissimilarity level (number of bases) used for the initial pseudo-single linkage algorithm (Huse et al., 2010). Except for the three lines in bold characters, all linear correlations were significant ($p < 0.05$). 87

Tableau 6-1 : Physico-chemical properties of soils..... 98

Tableau 6-2 : Initial PCP concentrations measured in the soil samples. Results are presented as the mean PCP concentration for each level, with the associated 95% confidence intervals (calculated whenever possible). 99

Tableau 6-3 : EC_{50} values extracted from the C-R curves presented in Figure 6-2. The numbers in brackets correspond to the 95% confidence intervals. 106

Tableau 6-4 : Ecotoxicological data used for SSD modeling. Considering that the dataset was dominated by worms and terrestrial plants, the SSD modeling process was carried out using the geometric mean of the EC_{50} available for each taxonomic group. All the ecotoxicological indicators (geometric mean and EC_{50}) are given in mg PCP/kg soil. The PAF (potentially affected fraction of species) corresponds to the cumulative proportion of species affected above their EC_{50} at a given PCP concentration (See the Material and methods section for the calculations). 107

LISTE DES FIGURES

Figure 1-1 : Exemple de RAD pour deux communautés différentes. Dans cet exemple, la richesse de la communauté 1 (~105 espèces) est plus faible que celle de la communauté 2 (~120 espèces). Aussi, comme l'abondance relative des espèces les plus abondantes est plus élevée pour la communauté 1, la dominance de cette communauté est supérieure à celle de la communauté 2. 6

Figure 1-2 : Comparaison de quatre distributions de probabilités couramment utilisées afin de construire les courbes concentration-réponse. Les différents modèles sont très semblables pour les réponses situées entre 20 et 80% mais diffèrent passablement pour les percentiles plus faibles et plus élevés. 11

Figure 1-3 : Réponse théorique d'un écosystème contaminé comparativement aux prédictions de la SSD. Tiré de Solomon et Takacs (2001). 16

Figure 1-4 : Nombre de données écotoxiques disponibles pour chaque contaminants répertoriées au Tableau 1-4. Les résultats sont présentés sous forme de pourcentage du nombre total de contaminants. 19

Figure 1-5 : Représentation schématique des régions conservées et hypervariables (V1 à V9) du gène 16S bactérien (tiré de http://www.alimetrics.net/en/index.php/dna-sequence-analysis). 23

Figure 1-6 : Principe général de l'amplification PCR. La première étape, la dénaturation, se réalise à haute température et permet de séparer l'ADN double-brin en 2 brins simples complémentaires. La deuxième étape, l'hybridation, permet aux amorces d'aller se « coller » à un endroit spécifique sur les brins d'ADN. Finalement, l'étape d'élongation, permet à la Taq polymérase de prolonger les brins d'ADN à partir de la fin des amorces. Le processus est répété entre 30 et 40 fois, chaque répétition permettant théoriquement de multiplier par 2 le nombre de brins d'ADN dans l'échantillon (tiré de http://users.ugent.be/~avierstr/principles/pcr.html). 25

Figure 1-7 : Gauche : Principe général du DGGE, tiré de Temmerman et al. (2004). Droite : Gel DGGE expérimental. 28

Figure 3-1 : Profil DGGE et courbe d'intensité lumineuse (en niveaux de gris) associée. Sur le graphique, la courbe visible à une intensité lumineuse de ~$4*10^6$ correspond à un profil DGGE où un échantillon PCR ne contenant pas d'ADN a été injecté. Afin de faciliter l'analyse, les images de gels DGGE sont redimensionnées en utilisant un facteur de 10 (10 240 pixels au lieu de 1024). Il a effectivement été observé ce redimensionnement facilitait la convergence du script d'optimisation décrit à la section 3.2. 39

Figure 3-2 : Délimitation entre le signal attribuable aux espèces abondantes (pics visibles) et celui attribuable aux espèces plus rares. La signification des chiffres 1, 2 et 3 est expliquée dans le texte plus bas. 40

Figure 3-3 : Résultats produits par l'algorithme de quantification des pics DGGE. À la gauche de la figure se trouve la portion de profil à analyser. L'information par rapport à la position des pics est déterminée directement à partir

du profil analysé et est entrée dans une boîte de dialogue (à droite de la figure). L'algorithme optimise alors les paramètres des distributions correspondant à chaque pic et présente les résultats sur un graphique. Les résultats peuvent être acceptés ou rejetés. 43

Figure 4-1 : Schematic representation of the elongation framework. The framework was developed with pyrosequencing datasets containing approximately 35 000 sequences per sample. The elongated RADs were therefore considered complete when the abundance of all the species summed to this value. The example is given for sample $BF_{100\%}$, and the results of the elongation process may be compared with the true RAD. Results are only presented for species rank under 1200 for readability but the actual richness of this sample was around 9500. 55

Figure 4-2 : Image of the gel synthesized from the datasets published by Nacke et al. (2011). 56

Figure 4-3 : Indicator values obtained using the parameters presented in Tableau 4-2, reported as the mean of all nine samples. Error bars correspond to the maximum and minimum values. All diversity measures were calculated using truncated RADs. 59

Figure 4-4 : PSRs extracted from the in silico DGGE profiles using the Matlab-based framework compared to true community richness at 98% similarity level. 61

Figure 5-1 : DGGE profiles of the 8 samples analyzed. The DNA loaded on the gel was extracted and amplified from a sandy loam (LM) or sandy soil (S). Samples were either clean (cln) or contaminated with pentachlorophenol (PCP) at concentrations ranging from 300 to 3,000 mg (kg dry wt soil)$^{-1}$. Left: 400 ng of PCR product per well. Right: 200 ng of PCR product per well. 77

Figure 5-2 : Diversity indices estimated using DGGE or Ion Torrent sequencing. The error bars correspond to the 95% confidence intervals that were calculated from the DGGE gel triplicates. DGGE diversity estimates were calculated from the RADs elongation framework presented in Lalande et al. (2013). Ion Torrent sequencing datasets were clustered at the 98% similarity level using Mothur (Schloss et al., 2009). 79

Figure 5-3 : Intensity of the DGGE band produced by loading different amounts of a PCR product amplified from the genomic DNA of *Escherichia coli*. All PCR products came from the same aliquot. Resulting intensities, calculated as the area under the peaks, are reported as a fraction of the intensity corresponding to a loading of 400 ng. 80

Figure 5-4 : The elongated DGGE RADs, each containing 35,000 reads, were randomly sampled without replacement at various depths ranging from 100 to 30,000 reads. The impact of sampling was quantitatively assessed by dividing the diversity indices calculated at a certain depth by the values calculated at a depth of 35,000. Since DGGE gels were run in triplicate, results were averaged for every sample. 82

Figure 5-5 : Proportion of the sequences deleted when applying an average quality filter (q > 20) over the Ion Torrent dataset. Upper: Deletion percentages for the different trimmed lengths. Lower: Observing that the sequences with lengths ranging between 133-138 and 157-163 bases contained more than half of all the sequences and were of particularly good quality, these sequences were extracted and ordered by increasing GC content, individually for each length. Results, reported individually for the two clusters (133-138 and 157-163), correspond to the average deletion percentage for the different lengths in each cluster. Error bars correspond to the 95% confidence intervals calculated for all the lengths in each cluster. 85

Figure 6-1 : DGGE migration profiles corresponding to the first replicate. DNA samples extracted from soil LM and soil S were subjected to PCR for the amplification of 16S rRNA gene sequences (the V3 variable region). The amplicons were separated on 8% polyacrylamide DGGE with a 32.5-72.5% denaturant gradient. The indices number in the sample names refer to the initial PCP concentrations in Tableau 6-2. 104

Figure 6-2 : Concentration-response (C-R) curves produced by the four indicators. The reduction percentages were calculated from the diversity or activity of the uncontaminated samples corresponding to a 0% reduction. **A.** Simpson's 1/D **B.** Shannon $\exp^{H'}$ **C.** Community richness **D.** FDA hydrolytic activity. For C-R modeling, the diversity indices calculated for the most contaminated samples were excluded from the data fitting process, since these diversity increases were associated with an important PCP toxicity not shown by DGGE profiles. The horizontal and vertical solid (soil LM) and dotted (soil S) lines show the position of the EC_{50} on the x-axis. ..105

Figure 6-3 : Species sensitivity distribution produced for PCP in terrestrial ecosystems. The SSD curve was drawn from the geometric mean of the EC_{50} available for the different species belonging to four taxonomic groups (Tableau 6-4). The black dots correspond to the EC_{50} of these 14 species.....108

LISTE DES SIGLES ET ABRÉVIATIONS

La liste des sigles et abréviations en ordre alphabétique :

ACV	Analyse du cycle de vie
ADN	Adénosine désoxyribonucléique
ARE	Analyse de risque environnemental
ARN	Adénosine ribonucléique
ASTM	American Society for Testing and Materials
DGGE	Électrophorèse sur gel en gradient dénaturant (*Denaturing Gradient Gel Electrophoresis)*
EC_x/ED_x	Concentration (dose) ayant un effet sur x% de la population (*Effect concentration*)
EQC	Critère de qualité environnementale (*Environmental Quality Criteria*)
FDA	Fluorescéine diacétate
HAP	Hydrocarbure aromatique polycyclique
HC_x	Concentration présentant un risque pour x% des espèces (*Hazardous concentration*)
LC_x/LD_x	Concentration (dose) mortelle pour x% de la population (*Lethal concentration*)
LOEC/LOEL	Plus faible concentration (dose) ayant un effet décelable (*Lowest observed effect concentration/limit*)
NOEC/NOEL	Plus forte concentration (dose) sans effet décelable (*No observed effect concentration/limit*)
NGS	Nouvelles technologies de séquençage d'ADN (*Next-generation Sequencing*)
OECD	Organisation de Coopération et de Développement Économique (*Organisation for Economic Co-operation and Development*)
OTU	Unité taxonomique opérationnelle (*Operational Taxonomic Unit*)

PAF	Fraction des espèces potentiellement affectées (*Potentially affected fraction of species*)
PCP	Pentachlorophénol
PCR	Réaction en chaîne par polymérase (*Polymerase Chain Reaction*)
PNUE	Programme des Nations Unies pour l'Environnement
RAD	Distribution de rang-abondance (*Rank-abundance distribution*)
RSSI	Index de stabilité relative des sols (*Relative Soil Stability Index*)
SIR	Respiration induite par l'ajout de substrat (*Substrate Induced Respiration*)
SSD	Distribution de sensibilité des espèces (*Species Sensitivity Distribution*)
Taq polymérase	Polymérase (enzyme) thermostable initialement isolée de la bactérie *Thermus aquaticus*

INTRODUCTION

L'importance de la diversité biologique sur la résistance et la résilience des écosystèmes face à des perturbations naturelles ou anthropiques n'est plus à démontrer (Cleland, 2012). Réalisant que la contamination des milieux naturels avait une influence négative sur la biodiversité des écosystèmes, plusieurs outils de gestion environnementale tels que l'analyse de risque environnemental (ARE) et l'analyse du cycle de vie (ACV) ont vu le jour. Afin d'être utilisés, ces outils nécessitent l'estimation de l'effet des contaminants sur la biodiversité des écosystèmes, qui repose majoritairement sur des essais de toxicité réalisés en laboratoire. Considérant que la capacité à estimer l'effet d'un contaminant sur un écosystème donné dépend notamment du nombre d'espèces ayant testées (Wheeler, Grist, Leung, Morritt, and Crane, 2002) ainsi que de la représentativité de ces espèces (Forbes and Calow, 2002), il est très préoccupant de constater qu'il existe peu de données écotoxiques pour les espèces peuplant les écosystèmes terrestres et que ces données sont fortement polarisées envers certaines espèces.

Les communautés microbiennes jouent un rôle de premier plan dans le bon fonctionnement des écosystèmes terrestres (van der Heijden, Bardgett, and van Straalen, 2008) et sont parmi les plus diversifiés sur Terre (Roesch et al., 2007). Constatant que ces communautés sont virtuellement absentes des bases de données écotoxiques terrestres (Sun, Pan, and Zhou, 2012; Suter II, Traas, and Posthuma, 2001), l'objectif principal de cette thèse est donc de développer une approche analytique permettant de quantifier la perte de diversité bactérienne causée par l'ajout de pentachlorophénol (PCP) à des échantillons de sol. La diversité bactérienne sera étudiée à l'aide de la technique de l'électrophorèse sur gel en gradient dénaturant (*Denaturing Gradient Gel Electrophoresis* ou DGGE). Bien que le DGGE soit actuellement incapable de produire des estimés de diversité corrélés à la « vraie diversité » bactérienne des sols (Blackwood, Hudleston, Zak, and Buyer, 2007), il a été rapporté que cette technique pourrait y arriver (Loisel et al., 2006). À ce jour, aucune approche analytique n'a été développée en ce sens.

Cette thèse présentera donc les différentes étapes de développement et de validation d'une telle approche. Cette dernière sera par la suite utilisée dans le but d'évaluer l'utilité d'une mesure de diversité comme indicateur de la toxicité du PCP sur les communautés bactériennes des sols. Le premier chapitre de cette thèse s'attarde sur la présentation et la critique des éléments ayant abouti à l'hypothèse de recherche. Cette hypothèse, les objectifs qui en découlent ainsi que la

méthodologie générale du projet sont présentés au chapitre 2. Par la suite, un aperçu général de l'approche analytique permettant de quantifier la diversité bactérienne des sols à partir de la technique du DGGE sera présenté (chapitre 3). Ce cadre de travail sera validé théoriquement (chapitre 4) et expérimentalement (chapitre 5). Le chapitre 6 présente l'application de ce cadre de travail afin d'évaluer l'effet du PCP sur la diversité bactérienne des sols. S'ensuivront une discussion générale, une présentation des contributions de cette thèse et, finalement, quelques conclusions et recommandations permettant d'orienter de futurs travaux de recherche.

CHAPITRE 1 REVUE DE LITTÉRATURE

La présente thèse étant centrée sur l'évaluation de l'effet du PCP sur la diversité biologique des bactéries dans le sol, il est logique de débuter cette revue sur des notions de diversité.

1.1 Diversité biologique dans les écosystèmes naturels

1.1.1 État et tendances

Selon l'Office québécois de la langue française, la diversité biologique, ou biodiversité, peut être définie comme le « nombre et l'abondance relative des gènes (diversité génétique), des espèces et des écosystèmes (des communautés) présents dans une zone donnée » (Office québécois de la langue française, 2012). De cette définition ressort l'idée selon laquelle la quantification de la biodiversité doit considérer à la fois le nombre d'espèces présentes (richesse de la communauté) mais également l'abondance relative des individus au sein des différentes espèces (dominance et équité). Ces éléments seront traités plus en détails subséquemment.

Vue sous l'angle du nombre d'espèces présentes, la biodiversité des écosystèmes naturels est astronomique. Tel que présenté dans le Tableau 1-1, plus de 1,75 millions d'espèces différentes ont été répertoriées uniquement pour le domaine des eucaryotes (Groombridge and Jenkins, 2002). Le nombre total d'espèces eucaryotes a été estimé par plusieurs auteurs et se situe généralement entre 5 et 10 millions (Baillie and Upham, 2013). En ce qui concerne les procaryotes (archées et bactéries), 10 000 espèces ont été décrites mais cela pourrait ne constituer qu'environ 0,1% de la diversité totale de ces deux domaines (Tamames and Rosselló-Móra, 2012), pour un total approximatif avoisinant les 10 millions. Il importe de mentionner que la notion d'espèce en biologie et en microbiologie est différente (Ereshefsky, 2010).

Tableau 1-1 : Nombre d'espèces décrit et estimé pour les trois domaines biologiques. Tiré de Groombridge et Jenkins (2002).

Domaine	Espèces décrites	Nombre total estimé
Archées	175	10 000 000[1]
Bactéries	10 000	
Eucaryotes[2]	1 750 000	14 000 000

[1] Tiré de Tamames et Rosselló-Móra (2012).

[2] Comprend les eucaryotes unicellulaires et multicellulaires.

Il va sans dire que cette diversité est une richesse à préserver. À l'aube du nouveau millénaire, sous l'égide du Programme des Nations Unies pour l'Environnement (PNUE), une vaste étude à laquelle plus de 1350 experts ont participé a été lancée; l'Évaluation des Écosystèmes pour le Millénaire. Cette étude visait à faire le point sur les changements survenus dans les écosystèmes ainsi que de proposer des pistes de solution afin de réagir à ces changements. Les conclusions de ce groupe de travail sont sans équivoque; induits par l'activité humaine, les changements de biodiversité des écosystèmes a été plus rapide dans les 50 dernières années qu'à n'importe quelle période de l'histoire humaine. Effectivement, le taux d'extinction des espèces actuellement observé dans les écosystèmes est plus de 100 fois supérieur aux taux passés (Millenium Ecosystem Assessment, 2005).

Les prédictions quant à l'évolution de la biodiversité dans les écosystèmes naturels ne sont pas encourageantes. Les facteurs ayant mené aux changements actuellement observés (modification des habitats, changements climatiques, invasion par des espèces étrangères, surexploitation et pollution) sont soit stable dans le temps, ne montrent aucun signe de ralentissement ou encore montrent des signes d'augmentation. Bien que très incertain, le taux d'extinction des espèces pourrait, dans un futur rapproché, être entre 10 et 100 fois supérieur au taux actuellement observé (Millenium Ecosystem Assessment, 2005). Aux États-Unis, environ le tiers des espèces végétales et animales est menacé d'extinction (McCann, 2000) et ces dernières pourraient disparaître.

1.1.2 Lien biodiversité – fonction – stabilité

Le déclin de la biodiversité des écosystèmes est une certitude (Butchart et al., 2010) et les prévisions ne sont guère plus encourageantes. En quoi cela constitue-t-il un problème? Philosophiquement parlant, savoir que l'Humanité est directement responsable de la disparition d'un nombre appréciable d'espèces vivantes peut constituer un problème éthique. Mais plus encore, en adoptant une vision des plus anthropocentriste, la recherche tend à démontrer qu'il existe un lien entre la biodiversité, les fonctions assurées par les écosystèmes (cycles biogéochimiques, production biotique, détoxification, etc.) et la résistance et/ou résilience des écosystèmes face à diverses perturbations. La résistance et résilience peut être définie comme la capacité d'un écosystème à maintenir ses fonctions suite à une perturbation (résistance) ainsi que sa capacité à récupérer suite à une perte de fonction (résilience). L'importance de la diversité sur la stabilité fonctionnelle des écosystèmes a été établi autant au niveau macroscopiques (Hooper et al., 2005) que microscopique (Deng, 2012; Girvan, Campbell, Killham, Prosser, and Glover, 2005). Par contre, le lien diversité-stabilité n'est pas direct. Incidemment, la stabilité d'un écosystème dépendrait d'une série de facteurs biotiques et abiotiques (Griffiths and Philippot, 2013). Globalement, la biodiversité promeut la stabilité en assurant une redondance fonctionnelle dans les écosystèmes. Puisque plusieurs espèces peuvent remplir le même rôle, la perte d'une espèce peut être compensée par une autre espèce (Cleland, 2012). De cette discussion, il est possible de déduire qu'il est très difficile de prédire la conséquence d'une perte de biodiversité sur les fonctions d'un écosystème. Ainsi, la perte d'une espèce clé peut, à elle seule, entraîner de graves conséquences. Bien que le lien quantitatif entre biodiversité, fonction et stabilité ne puisse être établi, une perte de diversité peut tout de même être considérée comme un risque pouvant à terme mener à une perte de fonction ou à l'effondrement de l'écosystème.

1.1.3 Quantification de la biodiversité

Jusqu'à maintenant, la biodiversité a été présentée d'un point de vue plutôt théorique. Par contre, afin de comparer la diversité de différents écosystèmes ou d'évaluer l'effet de différents traitements sur ces derniers, il importe d'être en mesure de quantifier la diversité. La diversité biologique a été divisée en trois éléments; les diversités *alpha, beta* et *gamma* (Whittaker, 1972). La diversité totale dans une région donnée (diversité *gamma*) a été définie à l'aide de deux éléments; la diversité évaluée à une échelle plus locale (diversité *alpha*) ainsi que les différences de diversité locales au

sein de la région globale (diversité *beta*). La présente thèse se concentrera uniquement sur la diversité *alpha*.

La quantification de la biodiversité passe généralement par la détermination de la distribution rang-abondance (*Rank-abundance distribution* ou RAD) représentant la communauté étudiée. La RAD (Figure 1-1) est une représentation graphique permettant d'observer l'évolution de l'abondance (relative ou absolue) des espèces composant une communauté en fonction du rang de celles-ci (le rang correspond simplement à la position d'une espèce donnée lorsque toutes les abondances sont triées en ordre décroissant).

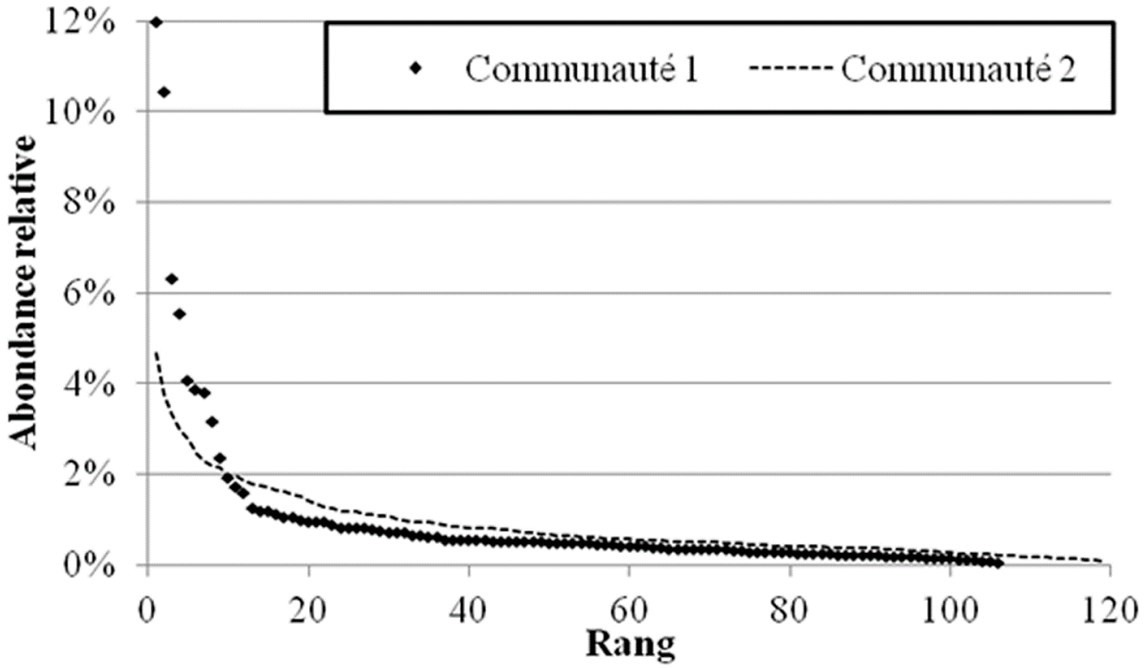

Figure 1-1 : Exemple de RAD pour deux communautés différentes. Dans cet exemple, la richesse de la communauté 1 (~105 espèces) est plus faible que celle de la communauté 2 (~120 espèces). Aussi, comme l'abondance relative des espèces les plus abondantes est plus élevée pour la communauté 1, la dominance de cette communauté est supérieure à celle de la communauté 2.

Tel que mentionné précédemment, la définition de diversité biologique fait implicitement référence à deux éléments : le nombre d'espèces (richesse) ainsi que l'abondance relative des différentes espèces (dominance/équité). La majorité des indices utilisés afin de quantifier la diversité d'une communauté (Tableau 1-2) tiennent compte de ces deux aspects. Bien que les équations permettant de calculer chaque indice de diversité soient passablement différentes, la majorité des indices sont

assez similaires. Il s'agit le plus souvent du temps d'une sommation réalisée sur la totalité des espèces répertoriées (S) tenant compte de l'abondance relative ou absolue de chaque espèce (p_i). Les indices ne tenant pas compte de ces deux éléments risquent d'être fortement polarisés envers une facette particulière de la diversité (par exemple, l'indice de Berger-Parker, Tableau 1-2). La valeur numérique que prendront les différents indices dépendra principalement du poids accordé aux espèces rares et aux espèces abondantes dans le calcul.

Tableau 1-2 : Liste non-exhaustive des indices de diversité utilisés afin de quantifier la diversité des communautés biotiques. S correspond au nombre d'espèces observées, p_i correspond à l'abondance relative de l'espèce de rang i et N correspond au nombre d'individus contenus dans l'échantillon.

Indice	Formule	Référence
Richesse (S)	S	(Blackwood et al., 2007)
Shannon (H')	$-\sum_{i=1}^{S} p_i * \ln(p_i)$	(Blackwood et al., 2007)
Shannon nb effectif	$exp^{(H')}$---	(Jost, 2006)
Simpson (D)	$\sum_{i=1}^{S} p_i^2$	(Magurran, 2004)
Simpson nb effectif (1/D)	$1/\sum_{i=1}^{S} p_i^2$	(Jost, 2006)
Berger-Parker (1/d)	$1/\max(p_i)$	(Blackwood et al., 2007)
Shannon équité (J')	$H'/\ln(S)$	(Blackwood et al., 2007)
Simpson équité ($E_{1/D}$)	$\frac{1/D}{S}$	(Blackwood et al., 2007)
Smith et Wilson équité (E_{var})	Voir [1]	(Magurran, 2004)

[1] $E_{var} = 1 - \frac{2}{\pi} arctan\left\{\sum_{i=1}^{S} \left(\sum_{i=1}^{S} \ln(p_i) - \sum_{j=1}^{S} \frac{\ln(p_j)}{S}\right)^2 / S\right\}$

La représentativité de ces indices dépendra donc de la capacité de la méthode d'échantillonnage utilisée à identifier toutes les espèces composant une communauté et à estimer le plus précisément

possible l'abondance de ces espèces. En pratique, toutes les espèces ne seront pas échantillonnées et l'abondance des espèces échantillonnées peut varier en fonction de plusieurs facteurs (grosseur des organismes, facilité de capture, méthode d'échantillonnage, etc.). Dans tous les cas, les indices calculés correspondent à une estimation de la diversité réelle des communautés.

Les indices les plus fréquemment utilisés dans la littérature sont la richesse, l'indice de Shannon (H') et, dans une moindre mesure, l'indice de Simpson (D ou 1/D). Ces trois indices accordent un poids bien différent aux espèces rares et abondantes. À un extrême, la richesse accorde exactement le même poids à toutes les espèces. L'indice sera donc fortement influencé par les espèces rares et accordera très peu de poids aux espèces abondances. L'autre extrême est représenté par l'indice de Simpson. En tant qu'indice de dominance, Simpson accorde beaucoup plus de poids aux espèces abondances et est très peu affecté par les espèces rares (et ultimement par la richesse de la communauté). L'indice de Shannon, exprimé sous sa forme $\exp^{H'}$, est en quelque sorte un intermédiaire entre la richesse et Simpson (Hill, Walsh, Harris, and Moffett, 2003).

Selon Jost (2006), tous les indices de diversité ne peuvent pas être utilisés afin de comparer quantitativement la diversité de deux communautés. Ce dernier soutient qu'afin de produire des résultats intuitifs, les indices doivent être transformés afin de représenter un nombre d'espèces effectif. Pour un indice donné, le nombre d'espèces effectif se détermine en évaluant quelle richesse permettrait d'obtenir exactement la même valeur numérique d'indice si toutes les espèces avaient la même abondance relative. Cela permet donc de comparer des communautés présentant à la fois une richesse et un patron de dominance différents. Dans le Tableau 1-2, la richesse (S), l'indice de Shannon ($\exp^{H'}$) ainsi que l'indice de Simpson (1/D) correspondent à des mesures du nombre effectif d'espèces.

1.2 Évaluation de l'effet des contaminants sur les populations et les écosystèmes

Selon Odum (1985), une des tendances pouvant être observées dans des écosystèmes stressés (i.e. contaminés) est une réduction de la diversité accompagnée d'une augmentation de la dominance. Les indices de diversité présentés précédemment constituent donc des outils intéressants afin de quantifier l'effet des contaminants sur les écosystèmes. Effectivement, en considérant simultanément la présence et l'abondance d'un grand nombre d'espèces, ces derniers tiennent implicitement compte des différentes interactions pouvant lier les espèces composant une communauté (prédation, milieu de support, etc.). Par contre, certaines difficultés viennent limiter leur utilisation en écotoxicologie. Effectivement, afin d'encadrer, de limiter ou d'interdire l'utilisation de certaines substances, de fixer des limites d'émission et de définir des critères de qualité environnementale (*Environmental Quality Criteria* ou EQC), il importe d'être en mesure de prédire les conséquences que pourraient avoir ces contaminants s'ils se retrouvaient dans les écosystèmes. Comme il est difficile et coûteux d'effectuer des essais de toxicité à une échelle pouvant servir de proxy pour un écosystème en entier, les tests de toxicité sont plutôt réalisés en laboratoire sur des populations (espèces uniques). Les données de toxicité concernant les différentes populations testées sont ensuite utilisées afin de prédire l'effet des contaminants sur les écosystèmes.

1.2.1 Essais de toxicité sur les populations

L'évaluation de l'effet d'un contaminant sur une population, c'est-à-dire un groupe d'organismes appartenant à la même espèce, se fait généralement à l'aide de tests réalisés en laboratoire. Ces essais sont pour la plupart normalisés et leur réalisation doit respecter certains critères stricts. Par exemple, la norme de l'*American Society for Testing and Materials* (ASTM) concernant les tests de toxicité ciblant le ver *Eisenia fetida* spécifie le nombre d'organismes devant être utilisés pour chaque réplica, l'âge et la taille requise des organismes, la durée du test, la composition du milieu de test (pH, dureté de l'eau, composition du sol) et plus encore (ASTM International, 2012).

L'effet des contaminants sur les populations est représenté sous la forme de courbes concentration-réponse. Plusieurs sous-échantillons du même milieu (sol, sédiment, eau) sont contaminés à différentes concentrations augmentant généralement en série géométrique. L'Organisation de

Coopération et de Développement Économique (OECD) recommande l'utilisation d'au moins 5 niveaux de contamination différents (OECD, 1992). En réalité, le facteur le plus important pour le choix du nombre et des niveaux de concentration utilisés est de s'assurer que les réponses obtenues (% d'effet) sont bien réparties entre 0% et 100%. Il faut idéalement obtenir au moins une concentration produisant un effet presque nul, une autre près de 100%, et quelques concentrations réparties entre 10 et 90%. De cette façon, il sera possible de s'assurer que la courbe concentration-réponse ajustée aux résultats est valide sur toute la gamme de pourcentages d'effet.

La courbe concentration-réponse proprement dite est obtenue en ajustant les paramètres d'une distribution de probabilités aux résultats des essais. Plusieurs formes de distributions sont couramment utilisées; lognormale, log-logistique, Weibull (types 1 et 2), etc. (Ritz, 2010). Tel que présenté à la Figure 1-2, ces modèles diffèrent surtout pour les percentiles faibles (< 20%) ou élevés (> 80%). Le choix du modèle à utiliser dépend principalement de la qualité de l'ajustement. Également, ces modèles se présentent sous plusieurs formes; à 2, 3 ou 4 paramètres (Ritz, 2010). Les modèles à 2 paramètres seront bornés entre 0% et 100%, et peuvent être utilisés lorsque les données sont présentées en termes de pourcentages d'effet. Sinon, les modèles à 3 (une seule borne minimale ou maximale) ou 4 (non-bornés) paramètres peuvent être favorisés.

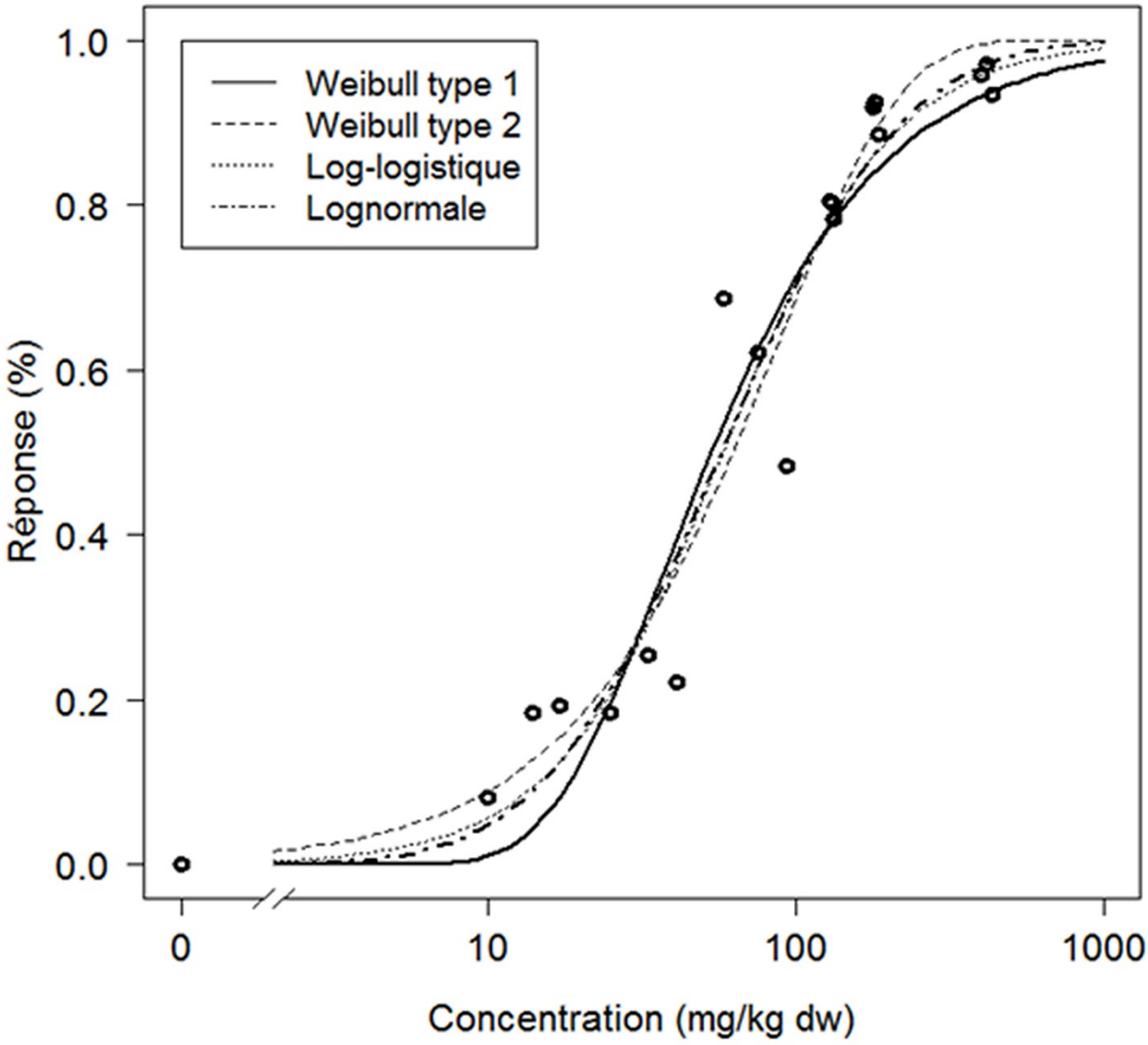

Figure 1-2 : Comparaison de quatre distributions de probabilités couramment utilisées afin de construire les courbes concentration-réponse. Les différents modèles sont très semblables pour les réponses situées entre 20 et 80% mais diffèrent passablement pour les percentiles plus faibles et plus élevés.

Plusieurs indicateurs toxicologiques peuvent être extraits directement des séries de données ou des courbes concentration-réponse. Les indicateurs les plus souvent utilisés sont les concentrations sans effets observables (*No Observed Effect Concentration* ou NOEC), les concentrations avec un faible effet observable (*Lowest Observed Effect Concentration* ou LOEC) et les concentrations avec effet sur un certain percentile (x) de la population testée (*Effect Concentration* ou EC_x).

La NOEC est la plus grande concentration testée ne présentant pas de différence statistique avec le contrôle tandis que la LOEC est la plus faible concentration testée qui présente une différence statistique avec le contrôle (Larsen and Hauschild, 2007). Ces deux indicateurs correspondent nécessairement à une concentration testée au laboratoire. Au contraire, les EC_x sont extraits de la courbe concentration-réponse et ne correspondent pas nécessairement à une concentration testée. Étant très dépendantes du choix des concentrations testées au laboratoire, il est reconnu que l'incertitude sur les NOEC et les LOEC est grande (Larsen and Hauschild, 2007). Certains auteurs recommandent même de ne plus produire/utiliser ces indicateurs (Warne and van Dam, 2008). Les EC_x, déterminées à partir de tous les points expérimentaux, sont moins incertains. Étant situé au centre de la distribution, l'indicateur le plus robuste est sans contredit l'EC_{50} (plages d'incertitudes plus étroites). De plus, les percentiles plus faibles ou plus élevés sont dépendants du choix d'une forme de distribution afin de tracer la courbe concentration-réponse (Figure 1-2).

Un élément central à la réalisation d'essais de toxicité est le choix d'une fonction suivie (*endpoint*). Cette fonction correspond au type de réponse ayant été mesurée; mortalité, reproduction, réduction de la taille, inhibition de la germination, etc. Le Tableau 1-3 présente les résultats obtenus par Roman et al. (2007). Ces derniers ont suivi simultanément plusieurs fonctions chez plusieurs espèces. Il existe des différences importantes entre les différents types d'indicateurs et les différentes fonctions suivies pour une même espèce. Au niveau des fonctions suivies, un facteur de 5 peut exister chez la même espèce (i.e. EC_{50} de *Chyronomus riparius* pour l'émergence et la survie).

Tableau 1-3 : Différences entre type d'indicateur et fonction suivie pour différents organismes exposés au cuivre. Adapté de Roman et al. (2007). Toutes les valeurs sont en mg de cuivre/kg de sol.

Espèce	NOEC	LOEC	EC_{10}	EC_{50}
Tubifex tubifex (28 jours)				
Survie	138	158	160 (136-190)	327 (302-354)
Croissance	78.3	102	43 (25-74)	126 (107-150)
Reproduction (jeunes/adultes)	78.3	102	79 (63-99)	98 (90-107)
Reproduction (cocons/adultes)	78.3	102	86 (72-101)	113 (105-121)
Hyalella azteca				
Survie (14-jours)	100	180	135 (103-168)	316 (281-355)
Croissance (28-jours)	53.2	95.4	75 (43-135)	194 (153-248)
Chironomus riparius				
Survie (14-jours)	< 180	180	150 (111-204)	320 (279-366)
Croissance (28-jours)	89.2	188	93 (65-127)	150 (125-169)
Émergence (28-jours)	59.5	89.2	33 (16-68)	59 (44- 80)
Lumbriculus variegatus (28 jours)				
Survie	114	140	126 (113-141)	211 (194-229)
Biomasse	80.5	103	70 (60-80)	126 (119-133)
Reproduction	80.5	103	97 (91-103)	105 (102-108)
Gammarus pulex (35 jours)				
Survie	94.7	176	73 (39-139)	151 (115-198)
Croissance	94.7	176	102 (60-169)	148 (132-186)

1.2.2 Évaluation de la sensibilité des écosystèmes

L'évaluation de la sensibilité des écosystèmes face aux contaminants se fait via l'utilisation des indicateurs toxicologiques disponibles pour les différentes populations pouvant évoluer dans l'écosystème considéré. La fraction des espèces potentiellement affectées (*Potentially Affected Fraction* ou PAF) est utilisée comme mesure de sensibilité et/ou stress toxique ressenti par un écosystème. Cet indicateur, généralement déterminé à partir d'une distribution de sensibilité des espèces (*Species Sensitivity Distribution* ou SSD), est notamment utilisé dans le but de développer des EQC pour les sols, les sédiments et l'eau dans plusieurs pays (Gaudet, Bright, Adare, and Potter, 2001; Scott-Fordsmand and Jensen, 2001; Sijm, van Wezel, and Crommentuijn, 2001; Stephan, 2001).

L'intégration de l'approche SSD en ARE fut la conséquence de maintes observations selon lesquelles la distribution de sensibilité de plusieurs espèces envers un contaminant avait des similitudes avec la distribution log-normale (Kooijman, 1987; Suter II, 2001; van Straalen and Dennenman, 1989; van Straalen and van Leeuwen, 2001). La courbe SSD est donc une représentation empirique de la variabilité des réponses toxiques observées chez diverses espèces composant un écosystème face à un contaminant. L'information véhiculée par cette courbe est purement statistique n'a pas d'interprétation biologique (Suter II et al., 2001). Dès lors, tout comme dans le cas de la modélisation concentration-réponse, aucune raison ne justifie l'utilisation d'une forme particulière de distribution (Aldenberg and Jaworska, 2000).

Bien qu'elle cherche à modéliser la sensibilité d'un écosystème face à un contaminant, les courbes SSD sont surtout utilisées afin de protéger les écosystèmes. Pour ce faire, l'indicateur le plus souvent utilisé pour la dérivation d'EQC est la concentration présentant un risque pour 5% des espèces (*Hazardous Concentration* ou HC_5). Plusieurs études ont montré que ce percentile de distribution avait la capacité de protéger les communautés aquatiques (Giddings, Solomon, and Maund, 2001; Hose and van den Brink, 2004; Leung, Grist, Morley, Morritt, and Crane, 2007; Maltby, Blake, Brock, and van den Brink, 2005; Roessink, Belgers, Crum, van den Brink, and Brock, 2006; Solomon, Giddings, and Maund, 2001; van den Brink, Brock, and Posthuma, 2001; Venturino, Montagna, and de D'Angelo, 2007; Wheeler, Leung, Morritt, Sorokin, and Rogers, 2002) et terrestres (Jensen, Smith, Krogh, Versteeg, and Temara, 2007; Semenzin, Temminghoff, and Marcomini, 2007).

Un élément primordial à considérer lors de la construction ou de l'interprétation d'une courbe SSD est la représentativité de la série de donnée écotoxique sous-jacente. Premièrement, tel que présenté au Tableau 1-3, il existe de grandes différences entre les différents types d'indicateurs et fonctions suivies chez une même espèce. Dès lors, il est préférable de construire les séries de données à partir d'essais similaires pour toutes les espèces. Puisque les espèces taxonomiquement très rapprochées sont généralement semblables en terme de sensibilité face à un contaminant (Boutin and Rogers, 2000; von der Ohe and Liess, 2004), il est préférable d'inclure dans ces séries de l'information sur une large gamme d'organismes. Par contre, dans plusieurs situations, les données disponibles représentent de façon disproportionnelle quelques espèces ciblées par certains organismes réglementaires (Frampton, Jänsch, Scott-Fordsmand, Römbke, and van den Brink, 2006). Si ces espèces sont particulièrement sensibles ou insensibles à un contaminant donné, la courbe SSD peut ne pas être représentative de l'écosystème modélisé.

La question du nombre de données requises afin d'estimer correctement les paramètres d'une courbe SSD est également un élément important à considérer. Tel que rapporté par Newman et al. (2001), bien que les agences réglementaires recommandent l'incorporation de 4 à 8 espèces dans le développement d'EQC, il apparait qu'entre 40 et 60 données seraient nécessaires pour que la prédiction soit juste. Cela dépasse largement le nombre de données disponibles dans la littérature pour essentiellement tous les contaminants. Une valeur de 10 semble représenter un bon compromis entre disponibilité et représentativité (Wheeler, Grist, et al., 2002).

L'approche SSD est donc utile afin de protéger les écosystèmes mais en pratique, elle présente quelques difficultés. Les conditions dans lesquelles les tests de laboratoire sont réalisés ne sont pas représentatives des conditions environnementales normales. Aussi, la disponibilité des données est souvent très problématique. Les séries de données contiennent donc trop peu d'indicateurs qui concernent souvent diverses fonctions suivies. Aussi, il arrive fréquemment que les espèces ciblées par les essais de toxicité ne soient pas représentatives de l'écosystème modélisé (i.e. utiliser des données sur des plantes potagères afin de modéliser l'effet d'un contaminant sur un écosystème forestier) (Forbes and Calow, 2002). De plus, tous les groupes taxonomiques ne sont pas représentés dans les bases de données écotoxiques. Finalement, la SSD simplifie énormément l'écosystème et ne tient pas compte des interactions entre les espèces. Ces interactions peuvent moduler à la hausse ou à la baisse la réponse d'un écosystème exposé à un agent stressant, particulièrement à de faibles et fortes concentrations (voir Figure 1-3) (Solomon and Takacs, 2001).

Contrairement à l'approche SSD, les indices de diversité tiennent implicitement compte des interactions entre les espèces.

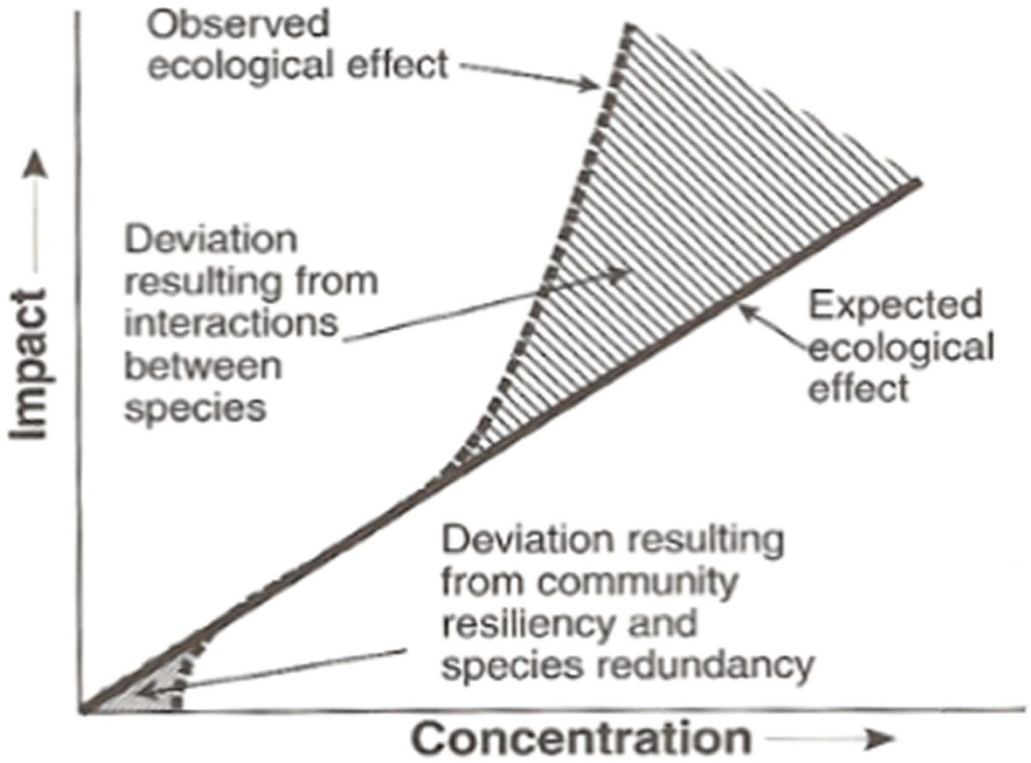

Figure 1-3 : Réponse théorique d'un écosystème contaminé comparativement aux prédictions de la SSD. Tiré de Solomon et Takacs (2001).

1.2.3 **Particularités des écosystèmes terrestres**

Les écosystèmes terrestres présentent plusieurs particularités qui complexifient ou limitent grandement l'application de l'approche SSD. Parmi ces éléments se trouve la question de la biodisponibilité des contaminants dans les sols. La biodisponibilité est la mesure de la proportion d'une contamination étant susceptible d'être ingérée, absorbée ou en contact avec les organismes. Elle module donc la toxicité d'une substance dans un environnement donné. Dans les sols, il est reconnu que la forme biodisponible d'un contaminant correspond à la portion dissoute dans l'eau interstitielle et non à celle adsorbée sur les particules de sol (Posthuma, Suter II, and Traas, 2001). Les contaminants très hydrophobes ont tendance à s'adsorber sur certaines particules composant les sols et donc à être peu biodisponibles. Dès lors, la composition d'un sol, notamment sa teneur en matière organique, est susceptible d'influencer la toxicité de ces contaminants. Par contre, la

biodisponibilité exprimée via la portion dissoute d'un contaminant n'est pas systématiquement apte à prédire la toxicité observée des contaminants sur certains récepteurs (Lock, De Schamphelaere, and Janssen, 2002). Étant donné la variabilité des propriétés physico-chimiques des sols, autant localement qu'à plus grande échelle, il est très probable que les sols utilisés au laboratoire afin de réaliser des essais de toxicité ne soient pas représentatifs de ceux retrouvés dans un écosystème donné. Bien qu'il soit théoriquement possible d'ajuster les données de toxicité afin de tenir compte de la biodisponibilité des contaminants dans différents types de sols, cet ajustement peut difficilement se faire en pratique étant donné que la composition du sol ayant été utilisé afin de réaliser un essai est rarement rapportée.

Une difficulté supplémentaire vient du fait que pour les organismes terrestres, la toxicité des contaminants s'exprime de deux façons différentes. Pour les organismes dits supérieurs (oiseaux et mammifères), l'exposition face à un contaminant se fait via plusieurs voies différentes (ingestion, inhalation, contact, etc.). Dès lors, la toxicité des contaminants doit être rapportée sous forme de doses quotidiennes ingérées (en mg de contaminant/kg de masse corporelle * jour). Pour les autres groupes taxonomiques (plantes, insectes, invertébrés, etc.), puisque la voie d'exposition face aux contaminants se fait majoritairement par contact direct, la toxicité est rapportée sous forme de concentration dans le média (sol ou eau). Afin d'intégrer les deux types d'indicateurs dans une même analyse, il faudrait réaliser une modélisation multimédias afin de faire le lien entre concentration dans le sol et dose ingérée par un certain récepteur. Cette pratique n'est pas courante. Le Tableau 1-4 présente la répartition des données écotoxiques contenues dans ECOTOX (U.S. Environmental Protection Agency, 2013), une des principales sources de données en libre-accès. Les données ont été traitées de façon à retirer les doublons. Chaque donnée répertoriée concerne donc un triplet contaminant – espèce testée – indicateur écotoxique unique.

Tableau 1-4 : Nombre de contaminants et types d'indicateurs couverts par la base de données ECOTOX (U.S. Environmental Protection Agency, 2013) pour les écosystèmes aquatiques et terrestres. Si la base de données contenait plus d'une entrée pour une même combinaison contaminant – organisme - type d'indicateur (i.e. une NOEC pour la reproduction et une NOEC pour la mortalité), cette combinaison a été comptabilisée une seule fois.

Écosystème	Type d'organismes	Nombre de contaminants	Nombre de données disponibles			
			NOEC	LOEC	EC_{50}	EC_x ($x \neq 50$)
Aquatique	Tous	5799	13253	9507	28389	22256
Terrestre	Mammifères et oiseaux	1897	3451	2982	2352	1669
	Contact direct (plantes, vers, insectes, etc.)	2322	15038	12917	4863	6026

Les données du Tableau 1-4 permettent de constater qu'il existe une différence assez marquée entre la disponibilité des données aquatiques et terrestres. Les données aquatiques sont non seulement plus nombreuses et couvrent près de trois fois plus de contaminants, mais elles correspondent également majoritairement à des EC_{50} et à d'autres EC_x, indicateurs jugés plus fiables que les NOEC et LOEC (Warne and van Dam, 2008). La Figure 1-4, obtenue en répertoriant le nombre d'espèces différentes testées pour chaque contaminant, montre clairement que peu de connaissances toxicologiques existent quant à l'effet des contaminants dans les écosystèmes et ce, peu importe le type de milieu. Effectivement, pour la majorité des contaminants, des données n'existent que pour 1 ou 2 espèces différentes. Pour les organismes terrestres à contact directs, les quelques 600 contaminant bien étudiés ($\geq$ 10 organismes testés) accaparent 90% de toutes les données disponibles. Pour l'aquatique, la répartition est semblable; 84% des données concernent 1200 contaminants bien étudiés.

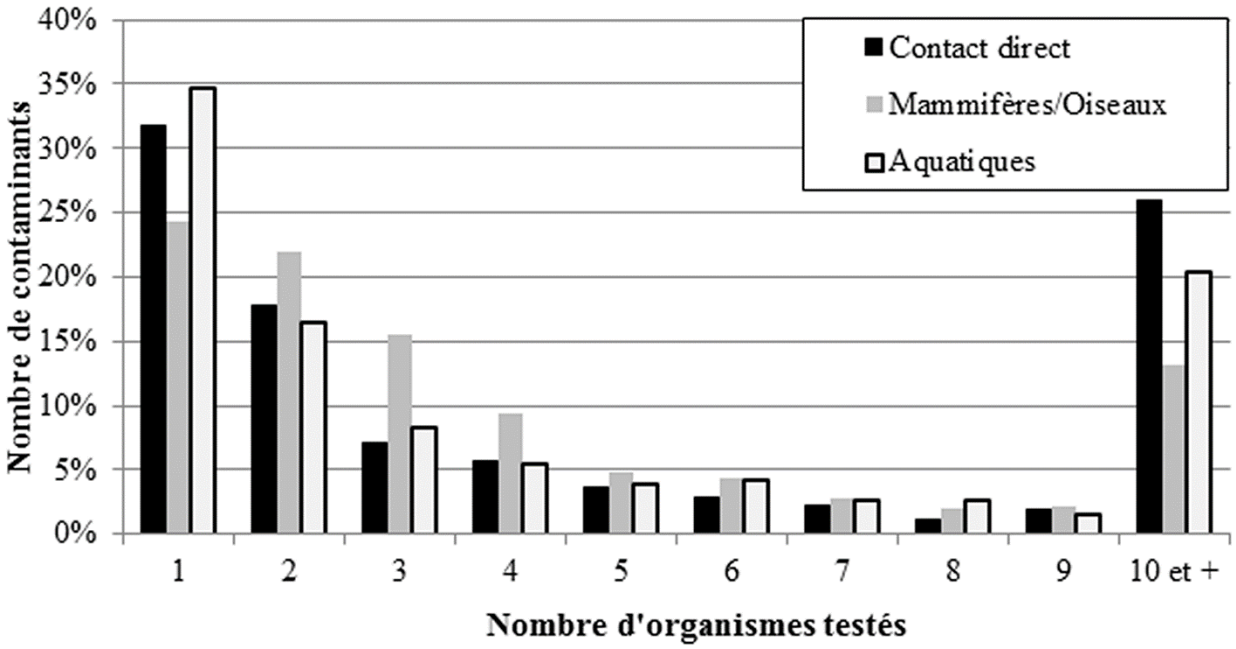

Figure 1-4 : Nombre de données écotoxiques disponibles pour chaque contaminants répertoriées au Tableau 1-4. Les résultats sont présentés sous forme de pourcentage du nombre total de contaminants.

Les données terrestres sont peu nombreuses mais également peu diversifiées en termes d'espèces testées. Les données disponibles sont fortement polarisées sur quelques espèces de plantes terrestres appartenant au phylum *Magnoliophyta* (laitue, maïs, soya, etc.) ainsi que sur certains vers et arthropodes. Tel que rapporté par d'autres auteurs, les microorganismes sont fortement sous-représentés dans les bases de données terrestres (Suter II et al., 2001). Pour ces organismes, les quelques données qui existent ont généralement été dérivées de mesures très générales d'activité enzymatique ou microbienne, de biomasse ou à l'aide de décomptes sur boîtes de Pétri. La faible disponibilité des données écotoxiques est donc une limitation importante à l'application de l'approche SDD pour les écosystèmes terrestres.

1.3 Importance, structure et diversité des communautés microbiennes des sols

La production d'indicateurs écotoxiques représentatifs des communautés microbiennes dans les sols permettrait la génération d'information hautement pertinente sur un groupe systématiquement sous-représenté dans les bases de données. En effet, ces communautés sont essentielles au bon fonctionnement des écosystèmes terrestres (pour un bon aperçu, voir Paul (2007) et van Elsas et

al. (2007)) et sont immensément diversifiées (Torsvik and Ovreas, 2007). En réalité, ces communautés sont parmi les plus diversifiées sur Terre; chaque gramme de sol peut contenir plus de 1000 espèces de champignons unicellulaires (Buée et al., 2009), 6000 espèces de bactéries (Roesch et al., 2007) et 300 espèces d'archées (Pires et al., 2012). Cette diversité génétique (et donc métabolique) permet aux communautés microbiennes d'être impliquées dans une multitude de processus permettant aux écosystèmes de bien fonctionner. Tel que revu par van der Heijden et al. (2008), les microorganismes du sol sont des contributeurs importants aux cycles du carbone, de l'azote et du phosphore dans les sols. Il a été estimé que cette communauté pourrait supporter entre 80 à 90% des réactions biochimiques se produisant dans les sols (Nannipieri et al., 2003). Les microorganismes du sol, via leurs capacités métaboliques diversifiées, sont également en mesure de dégrader une multitude de contaminants, participant ainsi grandement à la détoxification des milieux naturels.

Contrairement à la majorité des autres espèces terrestres, la contribution des microorganismes aux fonctions du sol peut difficilement être divisée entre les différentes espèces composant les communautés bactériennes, fongiques ou archéennes. La plupart de ces processus sont rendus possibles par l'action conjointe d'un grand nombre d'espèces. La discussion présentée à la section 1.1.2 (Lien biodiversité – fonction – stabilité) s'applique également aux communautés microbiennes. Le lien entre la diversité microbienne et la stabilité fonctionnelle des sols est d'ailleurs un sujet de recherche effervescent. Dans une étude, il a été expérimentalement montré que la diversité bactérienne avait un effet bénéfique à la fois sur la stabilité de la communauté et sur la capacité des fonctions du sol à récupérer à des perturbations (Girvan et al., 2005). Allant encore plus loin, Wittebolle et al. (2009) ont trouvé un lien entre l'équité initiale des communautés bactériennes et la stabilité du processus de dénitrification dans le sol. Encore ici, le lien entre diversité et stabilité n'est pas direct.

1.4 Évaluation de l'effet des contaminants sur la communauté microbienne d'un sol

La multitude d'indicateurs permettant d'étudier l'effet des contaminants sur les communautés microbiennes des sols ont fait l'objet de plusieurs revues (Imfeld and Vuilleumier, 2012; Kirk et al., 2004; Winding, Hund-Rinke, and Rutgers, 2005). De façon générale, ces techniques peuvent être divisées en deux grandes classes; les approches fonctionnelles et moléculaires.

1.4.1 **Approches fonctionnelles**

Les indicateurs de type fonctionnels sont principalement associés à l'étude des processus/réactions biochimiques rendus possibles par l'action des microorganismes. Cette classe de tests regroupe les mesures de respiration du sol, de décomposition de la matière organique, de biomasse microbienne, de potentiel d'activité enzymatique ou d'activité microbienne (Winding et al., 2005). Il s'agit donc de mesures très générales qui englobent dans un même indicateur l'activité des bactéries, des archées, des champignons et de tout autre organisme se trouvant ou s'étant trouvé dans le sol.

Bien que ces indicateurs soient très pertinents pour l'étude de la santé d'un sol (Killham and Staddon, 2002), leur utilisation en tant qu'indicateur de toxicité présente certaines limitations. Par exemple, dans une étude réalisée sur l'effet du plomb sur la respiration de 6 sols différents, aucune différence significative entre les contrôles et les échantillons contaminés (5 niveaux de contamination différents) ne fut observée (Lazzaro, Schulin, Widmer, and Frey, 2006). Dans une autre, il fut observé que ce même indicateur pouvait être multiplié par quatre suite à l'ajout de PCP (Scelza, Rao, and Gianfreda, 2008).

Puisque ces indicateurs sont très globaux, leur sensibilité aux différents contaminants peut être très variable. Ainsi, la biomasse microbienne s'est avérée assez sensible au PCP (Megharaj, Singleton, and McClure, 1998) mais pas au plomb (Liao, Chen, Zeng, and Huang, 2007). Parfois, ces indicateurs vont carrément mener à des conclusions erronées. Effectivement, si le contaminant est biodégradable, les quelques espèces de dégradeurs auront facilement accès à une source de carbone abondante et ces derniers augmenteront significativement en nombre et en masse. Cette augmentation, s'étalant sur deux ordres de grandeur dans le cas d'un sol contaminé au 2,4-D (Macur, Wheeler, Burr, and Inskeep, 2007), peut masquer la disparition d'autres espèces ne pouvant pas dégrader le contaminant.

Une autre limitation liée aux indicateurs fonctionnels provient du fait que ces derniers ne ciblent souvent pas les microorganismes directement. Ils ciblent un processus dont l'origine émane principalement de l'activité des microorganismes. Des facteurs autres que la toxicité des contaminants envers ces organismes peuvent donc influencer ces indicateurs. Par exemple, une étude réalisée sur l'effet des hydrocarbures aromatiques polycycliques (HAP) sur l'activité de quatre enzymes dans le sol a montré qu'un contaminant peut influencer l'activité enzymatique de deux façons différentes. Ainsi, à l'aide d'un index de stabilité relative des sols (*Relative Soil*

Stability Index ou RSSI (Becaert, Samson, and Deschenes, 2006)) basé sur l'activité enzymatique, il a été possible de voir que les contaminants ont un effet à la fois sur la communauté microbienne produisant les enzymes et sur les enzymes eux-mêmes (inhibition) (résultats personnels non publiés). Il est donc très difficile d'établir l'importance quantitative de chaque processus, ce qui est problématique lorsque l'objectif du test est de dériver une mesure de toxicité spécifique aux communautés microbiennes.

1.4.2 Approches moléculaires

1.4.2.1 La phylogénie microbienne à l'ère de la microbiologie moléculaire

Les approches moléculaires sont principalement basées sur l'étude de l'adénosine désoxyribonucléique (ADN) ou de l'adénosine ribonucléique (ARN) extrait du sol. L'émergence de ces techniques est intimement liée à ce qui a été dénommé la « Grande anomalie du décompte sur plaque » (traduction libre de *Great plate count anomaly*) (Staley and Konopka, 1985). Cette expression fait référence à la grande différence existant entre le nombre de microorganismes pouvant être cultivés en laboratoire sur des boites de Pétri et celui pouvant être observé directement au microscope. Cela a mené à la conclusion que la majorité des espèces microbiennes, jusqu'à 99% (Hugenholtz, 2002), requièrent des conditions de croissance différentes de celles généralement utilisées en laboratoire. Cela constituait un problème important pour le domaine de la phylogénie microbienne qui était principalement basé sur l'étude d'espèces isolées et cultivées au laboratoire. Depuis, le domaine de la phylogénie microbienne s'est transformé. À l'heure actuelle, la phylogénie procaryotique est largement basée sur l'étude de la composition du gène codant pour la petite unité ribosomale (le gène 16S) (Rajendhran and Gunasekaran, 2011). Ce gène, essentiel au bon fonctionnement de la cellule, sert de marqueur permettant de classer les procaryotes en espèces. En règle générale, la composition du gène 16S de deux organismes est comparée base à base et si la similitude entre les deux séquences est supérieure ou égale à 97%, il est admis que les deux organismes appartiennent à la même espèce (Stackebrandt and Goebel, 1994). Cette règle n'est pas universelle; la résolution du gène 16S n'est pas invariablement suffisante afin de diviser les procaryotes en espèces (Rosselló-Mora and Amann, 2001). Par contre, elle offre un cadre utile, fondé sur des bases scientifiques, permettant le classement des procaryotes en espèces. Il est important de mentionner que des marqueurs similaires existent pour les champignons dans les sols.

1.4.2.2 Extraction et amplification de l'ADN contenu dans le sol

Les approches moléculaires sont donc basées sur l'extraction et l'amplification en chaîne par la polymérase (*Polymerase Chain Reaction* ou PCR) de l'ADN contenu dans un sol (Oros-Sichler, Costa, Heuer, and Smalla, 2007). Les protocoles d'extraction permettent de lyser les cellules afin de récupérer et purifier leur matériel génétique. Cette procédure peut générer certaines distorsions (lyse plus efficace de certains organismes (Feinstein, Sul, and Blackwood, 2009) ou adsorption de l'ADN sur certaines particules de sol (Yankson and Steck, 2009)) et doit être optimisée en fonction du type de sol à l'étude. De plus, puisque l'ADN est une molécule pouvant persister dans le sol, la procédure d'extraction récupérera à la fois l'ADN appartenant aux cellules mortes et viables (Marschner, Yang, Lieberei, and Crowley, 2001).

Afin d'avoir suffisamment d'ADN pour permettre son analyse, il faut amplifier l'extrait de façon sélective. Tel que présenté à la Figure 1-5, le gène 16S bactérien se divise en plusieurs sections conservées (lignes minces) et sections hypervariables (V1 à V9). Les sections conservées correspondent à des régions probablement vitales au bon fonctionnement de la petite unité ribosomale ayant de fait subi très peu de mutations au fil du temps; elles sont donc communes à un grand nombre d'espèces appartenant à un même groupe phylogénétique. Les régions hypervariables changent d'une espèce à l'autre et peuvent donc servir de marqueur phylogénétique.

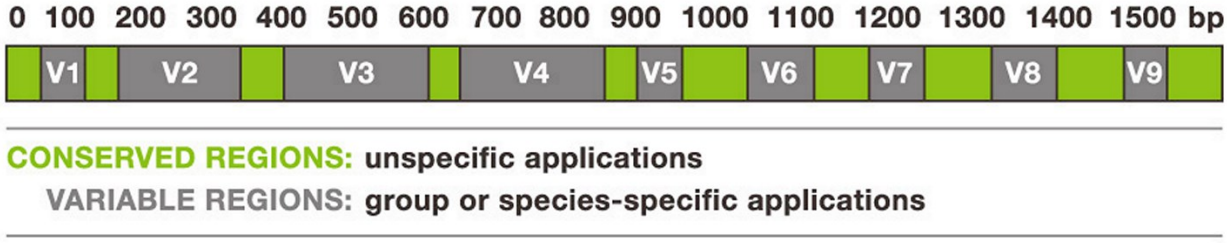

Figure 1-5 : Représentation schématique des régions conservées et hypervariables (V1 à V9) du gène 16S bactérien (tiré de http://www.alimetrics.net/en/index.php/dna-sequence-analysis).

L'amplification PCR se fait à partir de deux amorces (appelées « forward » et « reverse ») spécifiques aux organismes à l'étude. La séquence de nucléotides composant chaque amorce est complémentaire à une région conservée du gène à amplifier. Le choix des amorces est donc un facteur qui déterminera quelles espèces seront considérées lorsque le produit amplifié sera analysé. Les études de diversité utilisent généralement des amorces dites « universelles » qui ciblent simultanément un très grand nombre d'espèces. L'outil en ligne probeBase maintient une liste des

amorces pouvant être utilisées afin d'étudier les archées et les bactéries (Loy, Maixner, Wagner, and Horn, 2007). La spécificité des amorces peut être vérifiée à l'aide de TestPrime (Klindworth et al., 2013) ou Probe Match (Cole et al., 2009), qui vérifient quelles sont les espèces bactériennes ou archéennes contenues dans les bases de données qui seront ciblées par les deux amorces choisies. Ces outils en ligne permettent également de savoir quelle sera la longueur approximative des séquences amplifiées. Il est important de mentionner qu'il n'existe pas d'amorces permettant de cibler simultanément les communautés bactériennes, archéennes et fongiques constituant la communauté microbienne d'un sol. Les amorces offrent généralement une couverture élevée (% des espèces ciblées) pour un groupe donné et cibleront également quelques espèces provenant des autres groupes. Par contre, la longueur des fragments amplifiés chez les différents groupes sera généralement variable.

Tel que présenté à la Figure 1-6, l'amplification PCR est réalisée à partir d'un cycle de trois étapes réalisé entre 30 et 40 fois. Le premier cycle, réalisé à une température d'environ 95°C, permet de dénaturer (séparer) les deux brins complémentaires composant l'ADN. Par la suite, la température est réduite (la température dépend des amorces utilisées) de façon à permettre aux amorces d'aller s'apparier à la section qui leur est complémentaire sur les brins d'ADN dénaturés. Finalement, la température est ajustée à 72°C, la température optimale pour l'enzyme Taq (*Thermus aquaticus*) polymérase. Une réaction enzymatique complétera les brins d'ADN sur lesquels une amorce s'est fixée. Chaque cycle permet théoriquement de doubler le nombre de brins d'ADN contenus dans le milieu de réaction.

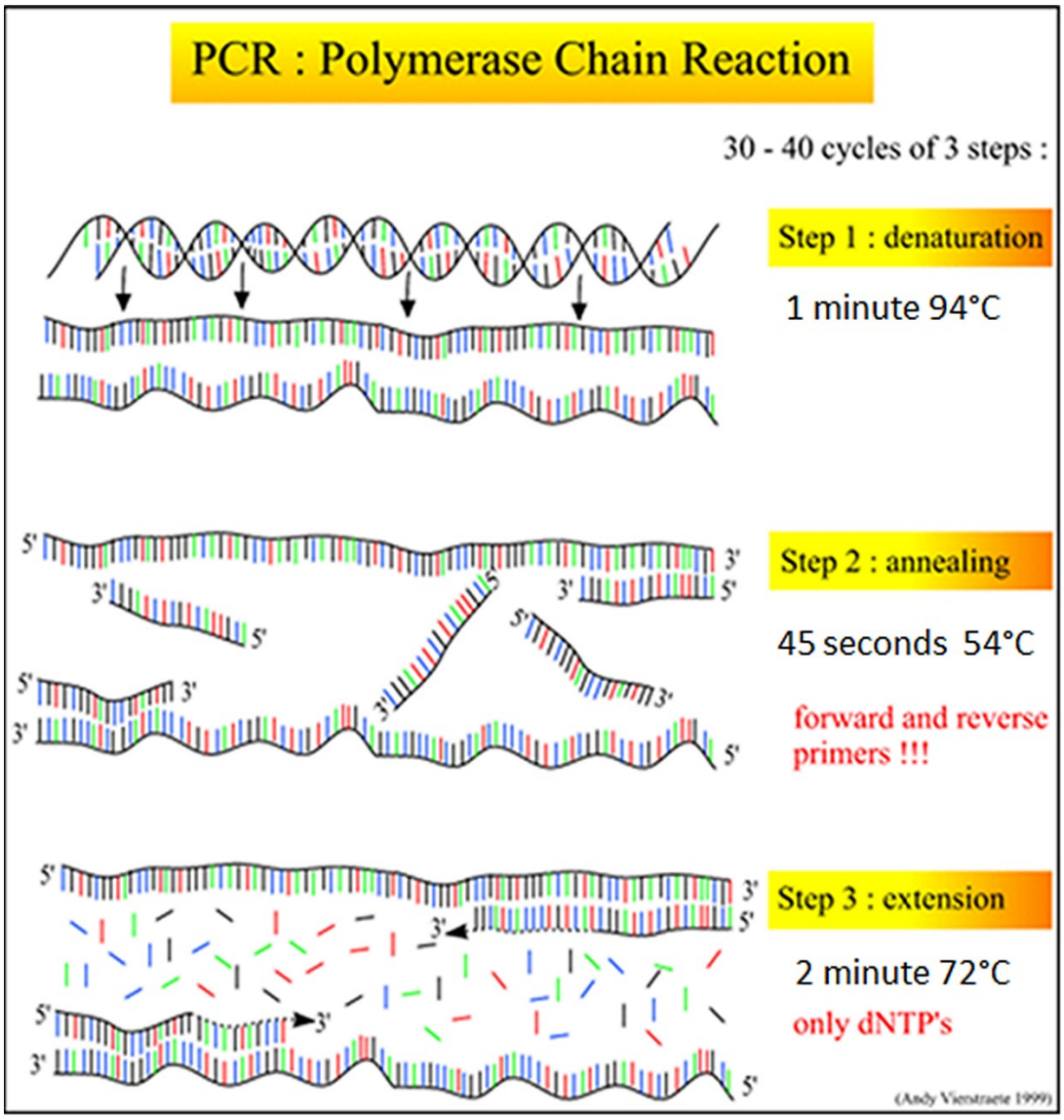
PCR : Polymerase Chain Reaction
30 - 40 cycles of 3 steps :
Step 1 : denaturation
1 minute 94°C
Step 2 : annealing
45 seconds 54°C
forward and reverse primers !!!
Step 3 : extension
2 minute 72°C
only dNTP's
(Andy Vierstraete 1999)

Figure 1-6 : Principe général de l'amplification PCR. La première étape, la dénaturation, se réalise à haute température et permet de séparer l'ADN double-brin en 2 brins simples complémentaires. La deuxième étape, l'hybridation, permet aux amorces d'aller se « coller » à un endroit spécifique sur les brins d'ADN. Finalement, l'étape d'élongation, permet à la Taq polymérase de prolonger les brins d'ADN à partir de la fin des amorces. Le processus est répété entre 30 et 40 fois, chaque répétition permettant théoriquement de multiplier par 2 le nombre de brins d'ADN dans l'échantillon (tiré de http://users.ugent.be/~avierstr/principles/pcr.html).

La distorsion introduite dans les études de diversité par la technique du PCR a été étudiée extensivement. Il a été rapporté que l'amplification PCR pourrait être plus ou moins efficace pour certaines séquences (et donc espèces) que pour d'autres, causant ainsi certaines distorsions dans

l'abondance relative des différentes espèces composant la communauté (Kurata et al., 2004; Polz and Cavanaugh, 1998). De plus, les polymérases utilisées afin de réaliser l'amplification possèdent un taux d'erreur intrinsèque (Cline, Braman, and Hogrefe, 1996). Ces erreurs, qui se produisent lorsque la polymérase substitue une base pour une autre lors de l'étape d'extension, seront amplifiées lors des cycles subséquents et créeront des séquences d'ADN qui n'existaient préalablement pas dans l'échantillon de base (augmentation de la diversité apparente). Une autre distorsion potentiellement importante peut se produire lors de l'amplification de l'ADN d'une communauté entière (grand nombre d'espèces différentes). Effectivement, le milieu de réaction contient un grand nombre de séquences d'ADN différentes qui peuvent, par hasard, être complémentaires sur une certaine région et s'apparier entre-elles. Cela mène à la formation de chimères qui peuvent, elles aussi, être amplifiées lors des cycles subséquents (Takahiro, 2003). Finalement, la qualité des amorces utilisées est également un aspect à considérer. La réaction de synthèse de ces molécules n'est pas parfaite; il a été rapporté que l'efficacité de réaction est de 99% pour chaque nucléotide ajouté (Life Technologies, 2013). Pour une amorce d'une longueur de 17 nucléotides, cela signifie que plus de 15% des amorces seront erronées (trop courtes ou encore contenant une substitution de base). Les amorces erronées ayant la capacité de s'apparier sur les brins d'ADN peuvent induire une variation dans la composition ou la longueur des séquences amplifiées.

Afin de réduire ces distorsions, les conditions d'amplification PCR doivent être soigneusement optimisées (Qiu et al., 2001; Takahiro, 2003). Tout d'abord, le choix d'une polymérase à haute-fidélité permet de réduire l'occurrence des erreurs d'amplification. Ensuite, l'utilisation d'une température d'appariement élevée permet d'augmenter la spécificité de l'appariement et de réduire la formation de produits non désirés. La spécificité de l'amplification peut être augmentée en réalisant un « touchdown PCR », c'est-à-dire de commencer l'amplification avec une température d'appariement élevée et de réduire cette dernière au fil des cycles (Korbie and Mattick, 2008). L'occurrence des chimères et autres erreurs d'amplification peut être réduite en abaissant le nombre de cycles d'amplification au minimum nécessaire (i.e. afin d'obtenir suffisamment d'ADN pour l'analyse). Finalement, plusieurs techniques de purification peuvent améliorer la qualité des amorces.

1.4.2.3 Analyse du produit amplifié à l'aide du DGGE

La technique du DGGE est une méthode de profilage permettant de séparer les brins d'ADN amplifiés sur la base de leur composition. Plusieurs autres techniques de profilage existent et il a été rapporté que ces dernières génèrent des résultats très similaires. Dès lors, le choix d'une ou l'autre repose le plus souvent sur l'expertise et l'équipement disponibles afin de réaliser une étude (Oros-Sichler et al., 2007). Le DGGE est sans contredit la plus utilisée des techniques de profilage (recherche Compendex).

Le DGGE se fait sur un gel d'acrylamide ayant un gradient linéaire d'urée et de formamide, deux agents permettant la dénaturation de l'ADN. Pour faire du DGGE, l'amorce « forward » doit contenir une pince GC (séquence de ~40 bases G ou C) (Sheffield, Cox, Lerman, and Myers, 1989). Dans l'ADN, le lien G-C est plus fort que le lien A-T. Tel que présenté à la Figure 1-7, cette pince GC se dénature difficilement et maintiendra les brins d'ADN ensembles lorsque ces derniers migreront vers le bas du gel. Le produit PCR est introduit dans un puits situé en haut du gel. Sous l'impulsion d'un courant électrique, les brins d'ADN migreront dans le gel et seront exposés à des conditions dénaturantes de plus en plus élevées. Petit à petit, les liaisons A-T et G-C se briseront et les brins d'ADN auront tendance à s'ouvrir. À une certaine position sur le gel, le brin est totalement ouvert (sauf la pince GC) et se retrouve immobilisé. La position d'un brin donné dépend de sa composition et comme chaque espèce possède une séquence d'ADN qui lui est propre, il est possible de séparer les séquences provenant de différentes espèces. L'analyse des gels DGGE considère implicitement, à tort, que chaque bande visible a été générée par une espèce unique et que l'intensité d'une bande est proportionnelle à l'abondance de l'espèce l'ayant produit. En permettant l'analyse d'environ 50 bandes, le DGGE ne génère de l'information que sur les espèces les plus abondantes de la communauté.

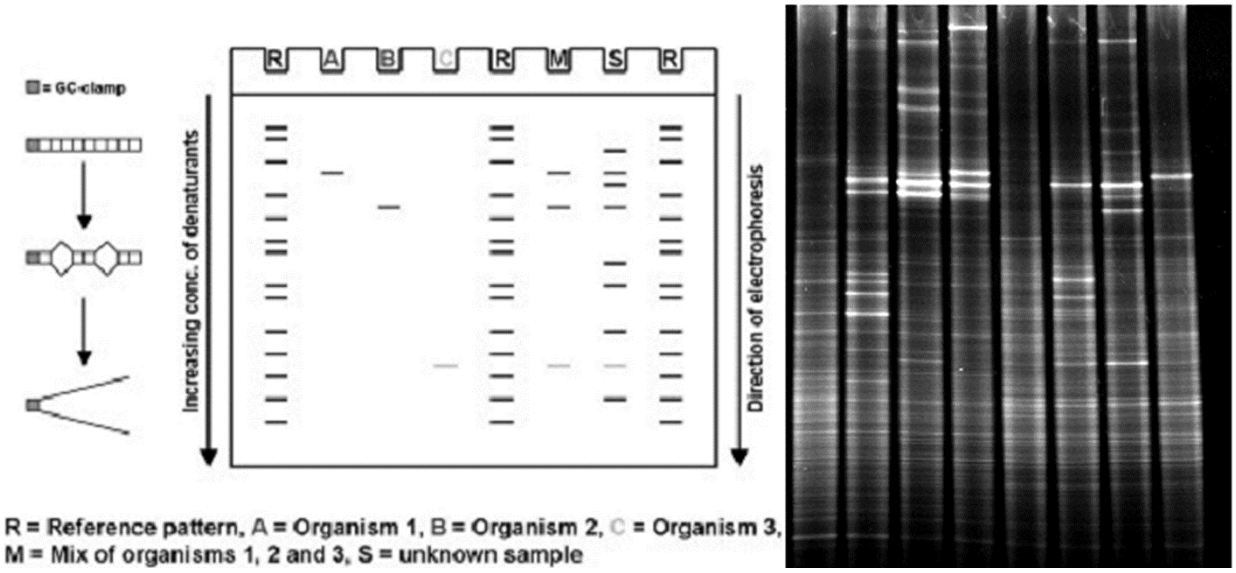
= GC-clamp
R A B C R M S R
Increasing conc. of denaturants
Direction of electrophoresis
R = Reference pattern, A = Organism 1, B = Organism 2, C = Organism 3,
M = Mix of organisms 1, 2 and 3, S = unknown sample

Figure 1-7 : Gauche : Principe général du DGGE, tiré de Temmerman et al. (2004). Droite : Gel DGGE expérimental.

Le DGGE a été maintes fois utilisé avec succès afin de différentier des communautés exposées ou non à différents contaminants. Cette approche s'est montrée efficace afin d'étudier des contaminations métalliques (Demanou et al., 2006; Li et al., 2006; Renella, Mench, Gelsomino, Landi, and Nannipieri, 2005) et organiques (Li, Zhang, Wu, Zhang, and Zhang, 2008; Macur et al., 2007; Zielezny, Groeneweg, Vereecken, and Tappe, 2006), pour ne citer que ces publications. Par contre, cet outil ne peut pas être considéré comme étant quantitatif. Effectivement, en considérant uniquement les espèces les plus abondantes, les méthodes de profilage sous-estiment la diversité des communautés étudiées (Hill et al., 2003; Narang and Dunbar, 2004). Plus encore, il a été démontré que ces estimés n'étaient pas corrélés à la diversité réelle des communautés (Blackwood et al., 2007) et ne peuvent donc pas être utilisés même à des fins comparatives. Par contre, à partir de communautés bactériennes simulées, un lien entre le bruit de fond en-dessous des bandes visibles sur les profils DGGE et la richesse de la communauté a été établi (Loisel et al., 2006). Ces mêmes auteurs ont par la suite breveté une approche permettant d'estimer l'indice de Simpson (D) à partir de profils DGGE (Loisel, Hamelin, Godon, Haegeman, and Harmand, 2009). Aucune approche permettant l'étude générale et quantitative de la diversité microbienne des sols à partir du DGGE n'a jusqu'à maintenant été développée.

Un élément fondamental qui n'a reçu aucune attention dans la littérature est la représentativité des patrons de dominance déterminés à partir des profils DGGE. À la base, le DGGE a été développé

en tant qu'outil de détection de mutations et peut théoriquement séparer deux séquences d'ADN qui diffèrent par une seule paire de bases (Myers, Maniatis, and Lerman, 1987; Sheffield et al., 1989). Par contre, en procédant au séquençage de l'ADN contenu dans des bandes DGGE, il a été clairement démontré que pour des communautés très diversifiées, les bandes DGGE contiennent une multitude de séquences d'ADN différentes (Schmalenberger and Tebbe, 2003). L'influence quantitative de cette co-migration sur les études de diversité basées sur le DGGE n'a jamais été évaluée.

La capacité du processus d'analyse des profils à estimer correctement l'abondance relative des espèces les plus abondantes n'a jamais été déterminée non plus. Les différents logiciels d'analyse des gels ont majoritairement été développés afin d'analyser des profils relativement simples. Dans un premier temps, le bruit de fond est soustrait, souvent à l'aide d'une bille virtuelle ayant un diamètre choisi par l'utilisateur qui roule sous le profil et qui soustraie le signal situé sous la bille. Par la suite, les bandes sont délimitées à l'aide de bornes linéaires manuellement positionnées ou encore à l'aide de distributions Gaussiennes. Si cette méthodologie peut donner de bons résultats pour des profils simples, il pourrait en être tout autrement pour les profils complexes tels que ceux produits par les communautés bactériennes dans les sols. Effectivement, à mesure que la diversité de la communauté augmente, le bruit de fond est de plus en plus complexe et important, et les bandes DGGE sont de plus en plus superposées les unes sur les autres. Les développeurs de logiciels admettent que ces profils peuvent être problématiques (communications personnelles) mais aucune étude n'a évalué la distorsion introduite par la méthode d'analyse dans les études de diversité.

1.4.2.4 Analyse du produit amplifié à l'aide des nouvelles technologies de séquençage d'ADN

Depuis quelques années, les nouvelles technologies de séquençage d'ADN (*Next-Generation Sequencing* ou NGS) ont fait leur apparition sur le marché. Au lieu de séparer les brins d'ADN sur la base de leur composition, ces appareils sont en mesure de « lire » les séquences. La plupart des plateformes NGS fonctionnent sur un principe général similaire. Sur une microplaque contenant des centaines de milliers de pores microscopiques, les brins d'ADN préalablement amplifiés par PCR sont réamplifiés. L'appareil introduit successivement dans les micropores chacune des bases qui constituent l'ADN (A, T, G et C). Lorsque la bonne base est ajoutée, la polymérase l'incorpore

à la chaine en extension. Le séquenceur est alors en mesure de mesurer un signal (émission lumineuse, d'ions, d'énergie, etc.) qui lui permet de déterminer quelle et combien de bases ont été ajoutées (l'intensité du signal mesuré est proportionnel au nombre de bases ajoutées). De cette façon, un très grand nombre de séquences ($> 10^6$) peuvent être lues pour chaque série d'analyse de l'appareil. En ajoutant aux amorces PCR une série de nucléotides propre à chaque échantillon, il est possible d'inclure plusieurs échantillons dans une même série d'analyse.

Par la suite, les séquences obtenue sont alignées et comparées base à base. À l'aide d'un algorithme de groupement, toutes les séquences similaires à plus d'un certain pourcentage, fréquemment fixé à 97%, sont regroupées en unités taxonomiques opérationnelles (*Operational Taxonomic Units* ou OTU). Les outils bio-informatiques nécessaires à l'analyse des séquences sont disponibles en ligne, notamment sur le site du *Ribosomal Database Project* (RDP) (Cole et al., 2009), ou encore via des logiciels libres (i.e. Mothur (Schloss et al., 2009)). Le nombre et l'abondance relative des OTUs peuvent servir à quantifier la diversité des communautés étudiées.

Il va sans dire que la puissance de ces approches dépasse largement la capacité des techniques de profilage. Les NGS ont connu une évolution fulgurante. Quelques revues, qui ont récemment été produites afin de comparer les différentes plateformes entre-elles, permettent de constater l'évolution de la capacité de ces appareils (nombre de séquences lues, longueur des séquences lues, coûts, etc.) (Glenn, 2011; Shokralla, Spall, Gibson, and Hajibabaei, 2012). Par contre, étant très récents comparativement au DGGE, la disponibilité des séquenceurs peut constituer une limitation pour plusieurs chercheurs et/ou laboratoires.

Bien que très prometteurs, les NGS doivent encore faire face à plusieurs défis techniques. Effectivement, sur certaines plateformes, les erreurs de séquençage sont encore relativement fréquentes (Loman et al., 2012; Quince, Lanzen, Davenport, and Turnbaugh, 2011). Ces erreurs causent une surestimation de la richesse des communautés étudiées (Kunin, Engelbrektson, Ochman, and Hugenholtz, 2010). Certains outils bio-informatiques ont néanmoins été développés afin d'identifier et corriger les erreurs de séquençage (Huse, Welch, Morrison, and Sogin, 2010; Quince et al., 2011). En plus des erreurs, quelques auteurs ont évoqué que la capacité des NGS à produire des séquences de bonne qualité pourrait varier en fonction de la composition des séquences. Ainsi, les séquences riches en G ou en C posent généralement problème (Jaenicke et al., 2011). Il a même été rapporté que cette problématique pourrait causer des distorsions intra-échantillons dans l'abondance relative des OTUs (Pinto and Raskin, 2012). Aussi, dépendamment

de l'indice de diversité utilisé afin de comparer les communautés, il sera nécessaire d'atteindre une profondeur de séquençage plus ou moins élevés (nombre de séquences de bonne qualité par échantillon). Ainsi, l'estimation de la richesse d'une communauté requiert la lecture de beaucoup plus de séquences que d'autres indices (i.e. Shannon et Simpson) (Pinto and Raskin, 2012).

1.5 Le pentachlorophénol comme cadre d'étude

Le PCP constitue un cadre d'étude très intéressant et d'actualité. Ayant été très utilisé afin de traiter le bois et blanchir le papier, ce dernier se retrouve fréquemment dans les écosystèmes naturels canadiens (Canadian Council of Ministers of the Environment, 1999). Le PCP est un contaminant hydrophobe ayant tendance à s'adsorber sur la matière organique (Puglisi et al., 2009). De plus, ce composé est ionisable (pKa = 4.7) (Boyle, 2006), ce qui implique que sa solubilité aqueuse augmente avec le pH (Arcand, Hawari, and Guiot, 1995). Dès lors, la biodisponibilité et la toxicité du PCP peut varier en fonction des propriétés physico-chimiques des sols.

En plus d'avoir des propriétés intéressantes, le PCP est un contaminant dont la toxicité envers divers organismes terrestres a été testée. Cette substance semble être particulièrement toxique envers les plantes terrestres. Effectivement, la valeur seuil (concentration au-delà de laquelle un effet peut commencer à être ressenti) du PCP a été déterminée à 5 mg/kg pour les plantes terrestres alors qu'elle est de 31 mg/kg pour les invertébrés (U.S. Environmental Protection Agency, 2007). Sur la base des données toxiques et écotoxiques disponibles, l'EQC du PCP dans les sols canadiens a été fixé à 7.6 mg/kg (Canadian Council of Ministers of the Environment, 1999). Une valeur semblable de 12 mg/kg est en vigueur aux Pays-Bas (Ministry of Housing Spatial Planning and the Environment, 2009).

L'effet du PCP sur les microorganismes dans des sols a également fait l'objet de quelques études. Dans une étude, il a été montré que des concentrations de PCP entre 10 et 50 mg/kg avaient un effet négatif significatif sur la croissance de *Phanerochaete chrysosporium*, un champignon largement étudié pour ses capacités à dégrader les xénobiotiques (Leštan, Leštan, Chapelle, and Lamar, 1996). Ces auteurs ont également établi qu'il existait une forte corrélation entre la biomasse microbienne (de *P. chrysosporium*) et une mesure simple d'activité enzymatique, la mesure du potentiel d'activité hydrolytique de la fluorescéine diacétate (FDA). Scelza et al. (2008) ont comparé plusieurs paramètres biochimiques de sols agricoles propres et artificiellement contaminés au PCP (50 mg/kg). Parmi ces paramètres, l'activité de plusieurs enzymes

(déhydrogénase, arylsulfatase, phosphatase et uréase), ainsi que les mesures de biomasse microbienne, ont été significativement affectées par la présence du PCP. L'activité de la glucosidase, une enzyme impliquée dans le cycle du carbone, n'a pas été influencée par le PCP. Quant à elle, la respiration du sol a été significativement affectée par le PCP, mais à la hausse. Ces résultats sont similaires à ceux obtenus par Megharaj et al. (1998) pour des sols contaminés « naturellement », situés à proximité d'une usine de traitement du bois (mesures de biomasse microbienne et d'activité enzymatique). Dans une autre étude, Martí et al. (2011) ont étudié l'effet du PCP sur la respiration cumulative et la respiration induite par l'ajout de substrat (*Substrate Induced Respiration* ou SIR) de deux sols différents contaminés à plusieurs concentrations différentes. À l'aide d'une modélisation concentration-réponse, ces auteurs ont généré des paramètres écotoxiques (EC_{20}, EC_{50} et LOEC) pour le PCP. Ces auteurs ont également produit des gels DGGE pour les différents échantillons de sols. Par contre, le seul paramètre écotoxique déterminé à partir des DGGE est la LOEC, qui fut déterminée visuellement. Les LOEC relatifs à la SIR furent inférieurs à ceux déterminés à partir des DGGE. Par contre, il est important de mentionner que les concentrations en PCP n'ont pas été mesurées dans les sols, ce qui limite l'interprétation des résultats.

CHAPITRE 2 PRÉSENTATION GÉNÉRALE DES TRAVAUX

2.1 Problématique

La revue précédente a permis de mettre en évidence le rôle crucial joué par la diversité biologique (autant microscopique que macroscopique) dans la capacité des écosystèmes naturels à fonctionner à long terme, malgré les différentes perturbations naturelles et anthropiques que ces derniers doivent subir. En ayant un effet négatif sur la biodiversité, la contamination des écosystèmes naturels risque également de réduire la capacité de ces derniers à supporter de futures perturbations. Afin d'éviter que les activités humaines aient des effets inacceptables sur les écosystèmes, différents outils tels que l'ARE et l'ACV ont été développés. Une limitation importante à l'application de ces outils est la disponibilité des données écotoxiques, qui font particulièrement défaut pour les écosystèmes terrestres.

Considérant qu'ils sont virtuellement absents des bases de données écotoxiques, la génération de données de toxicité spécifiques aux microorganismes du sol constitue un moyen permettant de s'attaquer à cette problématique. Vu l'importance de ces communautés dans les écosystèmes, tant du point de vue de leur diversité que de celui de leur importance fonctionnelle, ces données constitueraient de l'information hautement pertinente afin de mieux comprendre l'effet des contaminants en milieux terrestres.

Pour ce faire, plusieurs indicateurs pourraient être utilisés. Parmi ceux-ci, une mesure de toxicité basée sur la diversité génétique des communautés microbiennes dans les sols présente plusieurs avantages. Effectivement, en tenant compte de l'abondance relative des différentes espèces composant les communautés étudiées, une mesure basée sur la diversité microbienne a peu de chances d'augmenter suite à une contamination, même si le contaminant est facilement biodégradable. De plus, en divisant la communauté microbienne en plusieurs composantes distinctes (les espèces) plutôt qu'en un tout homogène, les mesures de diversité pourraient être plus sensibles à une contamination que les mesures fonctionnelles. Finalement, cet indicateur correspond à ce que la modélisation SSD cherche à reproduire (réponse de communauté face à un stress toxique), mesuré à l'échelle microscopique plutôt qu'à l'échelle macroscopique.

Par contre, l'utilisation de la diversité microbienne en tant qu'indicateur de toxicité présente également des difficultés. Premièrement, les approches moléculaires ne permettent pas la

considération simultanée des différents groupes composant les communautés microbiennes. Le domaine des bactéries, présentant généralement la plus grande diversité génétique dans les sols, est généralement sélectionné. La limitation la plus importante à l'utilisation d'une mesure basée sur la diversité bactérienne des sols est sans contredit la capacité des techniques actuelles à produire rapidement et à faible coûts un estimé quantitatif de diversité. Les plateformes NGS possèdent cette capacité mais demeurent pour le moment peu disponibles pour plusieurs chercheurs et/ou laboratoires. Au contraire, la technique du DGGE est très répandue et permet d'analyser rapidement un grand nombre d'échantillons. Par contre, aucune approche analytique ne permet actuellement d'utiliser le DGGE à des fins quantitatives.

Afin de développer une telle approche, le PCP semble être une substance idéale. En effet, ce contaminant se retrouve fréquemment dans les écosystèmes naturels canadiens, possède des propriétés physico-chimiques susceptibles de faire varier sa toxicité en fonction de la composition du sol utilisé et a été testé sur plusieurs espèces terrestres différentes. De plus, son effet sur les microorganismes du sol a fait l'objet de quelques publications. Il sera intéressant de comparer la réponse et la sensibilité d'un indicateur basé sur la diversité bactérienne des sols à des mesures biochimiques. À cette fin, la mesure du potentiel d'activité hydrolytique de la FDA, un essai simple produisant des résultats corrélés aux mesures de biomasse microbienne, semble être idéale. Effectivement, la biomasse microbienne s'est avérée être un indicateur sensible au PCP dans plusieurs études différentes.

2.2 Hypothèse de recherche et objectifs

L'hypothèse de recherche qui sous-tend les travaux présentés dans cette thèse est la suivante :

> *La technique de l'électrophorèse sur gel en gradient dénaturant peut permettre la production d'un indicateur écotoxique quantitativement représentatif de l'effet du pentachlorophénol sur la diversité taxonomique des bactéries du sol.*

L'objectif principal de cette thèse est donc de développer une approche analytique permettant de quantifier la perte de diversité bactérienne causée par l'ajout de PCP à des échantillons de sol à partir de la technique du DGGE. Puisque l'effet du PCP est déterminé en comparant la diversité

d'un échantillon propre à celle d'un échantillon contaminé, cette approche analytique devra être en mesure de produire des estimés de diversité corrélés à la diversité des communautés étudiées.

Les objectifs spécifiques de ce projet sont :

1. Proposer une approche analytique permettant d'extraire des profils DGGE de l'information quantitative sur la dominance et la richesse des communautés étudiées;
2. Développer, à partir de profils DGGE *in silico*, un modèle empirique permettant de produire des estimés de diversité corrélés à la diversité des communautés étudiées à partir de l'information extraite des profils DGGE;
3. Valider ce modèle sur des échantillons réels en comparant la diversité estimée à partir du DGGE à celle produite par une plateforme NGS;
4. Quantifier l'effet du PCP sur la diversité bactérienne des sols et comparer cette réponse à celle basée sur la mesure du potentiel d'activité hydrolytique de la FDA des sols.

2.3 Méthodologie générale

Afin d'atteindre les objectifs, il est proposé de :

i. Développer une approche analytique ainsi que les outils logiciels associés permettant d'extraire des profils DGGE de l'information pertinente et quantitative sur les communautés bactériennes étudiées. (chapitre 3)

ii. À partir de profils DGGE *in silico*, évaluer la performance de divers logiciels d'analyse à produire un portrait réaliste des communautés bactériennes étudiées. (chapitre 4)

iii. À partir des mêmes profils *in silico*, développer et paramétrer une approche empirique permettant de quantifier la diversité bactérienne des sols à partir de la technique du DGGE. (chapitre 4)

iv. Valider cette approche sur des échantillons réels en comparant la diversité bactérienne estimée à l'aide du DGGE à celle générée par une plateforme NGS. (chapitre 5)

v. Réaliser des essais de toxicité sur le PCP en ciblant les communautés microbiennes dans les sols. Pour ce faire, évaluer l'effet du PCP sur la diversité des communautés bactériennes ainsi que sur le potentiel d'activité hydrolytique de la FDA des sols. (chapitre 6)

vi. Rassembler un maximum de données écotoxiques sur le PCP qui couvriront la plus large gamme possible d'espèces terrestres (plantes, vers, insectes, etc.). (chapitre 6)

vii. Réaliser une modélisation concentration-réponse sur les résultats des essais de toxicité et produire une distribution de sensibilité des espèces (courbe SSD) à partir des données écotoxiques concernant les espèces terrestres. (chapitre 6)

viii. Extraire des courbes concentration-réponse et SSD certains indicateurs écotoxiques fréquemment utilisés dans la littérature et comparer la sensibilité des écosystèmes terrestres au PCP à l'échelle microscopique et macroscopique. (chapitre 6)

2.4 Présentation des manuscrits

Le premier manuscrit (chapitre 4) présente les étapes de développement de la méthodologie permettant de quantifier la diversité bactérienne dans les sols à partir de la technique du DGGE. Les travaux présentés dans ce manuscrit se sont basés sur l'analyse de profils DGGE *in silico* construits à partir des résultats d'une plateforme NGS. En travaillant avec des communautés bactériennes connues, il a été possible de comparer les profils de dominance produits par plusieurs logiciels d'analyse de gels DGGE aux profils réels caractérisant ces communautés. Cela a également donné l'opportunité d'étudier l'occurrence et l'influence des épisodes de co-migration des séquences d'ADN sur les estimés de diversité produits à l'aide de cette technique. Reconnaissant la supériorité d'une approche analytique basée sur le logiciel Matlab développée dans le cadre de ce projet, une méthode empirique permettant de compléter les RAD partielles issues de la quantification des bandes DGGE a été développée et calibrée. Ce modèle d'élongation se base sur le ratio pics-sur-signal (*Peaks-to-signal ratio* ou PSR), un paramètre estimé suite à l'étape de soustraction du bruit de fond des profils DGGE. Cette approche a permis d'estimer avec une très bonne précision la diversité bactérienne réelle des communautés étudiées. Ce manuscrit, intitulé « *A new framework to accurately quantify soil bacterial community diversity from DGGE* », a été publié dans la revue *Microbial Ecology*.

Dans le but de valider la méthode d'élongation des RAD DGGE, le second manuscrit (chapitre 5) a comparé les estimés de diversité produits par le DGGE et par la plateforme NGS Ion Torrent PGM. La comparaison de ces deux techniques a permis d'identifier des distorsions dans les résultats produits par les deux approches. Pour le DGGE, il a été démontré que la quantité d'ADN

normalement injectée dans les puits pouvait causer une saturation locale du gel. Ce phénomène, qui se produit dans les régions du gel où une grande quantité d'ADN migre, réduit artificiellement l'abondance estimée des espèces les plus dominantes d'une communauté. Pour l'Ion Torrent, les multiples sources de distorsions furent majoritairement reliées à la qualité des séquences produites; erreurs de séquençage fréquentes, trop peu de séquences par échantillon et faible qualité des séquences riches en GC. Une fois ces distorsions partiellement corrigées, il a été possible de montrer que la diversité estimée à l'aide des deux approches étaient fortement corrélée. Dès lors, utilisées à des fins comparatives, l'une ou l'autre de ces techniques générerait des résultats similaires. Ce manuscrit, intitulé « *Can DGGE and Ion Torrent sequencing yield similar quantitative conclusions when comparing the diversity of soil bacterial communities?* », a été soumis à la revue *FEMS Microbiology Ecology*.

Le troisième manuscrit (chapitre 6) a cherché à évaluer l'utilité des indices de diversité afin de quantifier l'effet du PCP sur les communautés bactériennes des sols. La réponse obtenue à l'aide d'indices de diversité fut comparée à celle produite par des mesures de potentiel d'activité hydrolytique de la FDA. Cet essai fut sélectionné puisqu'il produit des résultats corrélés aux mesures de biomasse microbienne, un indicateur s'étant avéré sensible au PCP. Les indices de diversité furent plus de 25 fois plus sensibles au PCP que la mesure du potentiel d'activité de la FDA. De plus, il n'a pas été possible de différentier les courbes concentration-réponse produites à partir des indices de Simpson (1/D) ou Shannon ($\exp^{H'}$) et la SSD représentative de l'effet du PCP sur les écosystèmes terrestres pour tous les percentiles de distribution situés entre 10 et 50%. Ce manuscrit, intitulé « *Assessing the impact of pentachlorophenol on soil bacterial communities using ecological diversity indices* », a été soumis à la revue *The ISME Journal*.

CHAPITRE 3 DÉVELOPPEMENT D'UNE APPROCHE ANALYTIQUE PERMETTANT L'EXTRACTION D'INFORMATIONS QUANTITATIVES D'UN PROFIL DGGE

Les logiciels permettant d'analyser les profils DGGE présentent plusieurs limitations quand vient le temps de travailler avec les communautés bactériennes des sols. Étant très diversifiées, les profils correspondant à ces communautés sont caractérisées par un important bruit de fond ainsi que d'un grand nombre et d'une forte superposition des bandes. Dès lors, des outils informatiques mieux adaptés à ces profils complexes furent développés à l'aide du logiciel Matlab.

Les outils présentés dans cette section serviront à extraire des profils DGGE l'information la plus juste possible quant à l'abondance relative des espèces dominantes des communautés bactériennes étudiées. De plus, la méthode de soustraction du bruit de fond présentée ci bas cherchera à permettre l'extraction de ratios pics-sur-signal aussi représentatifs que possible. L'approche globale présentée dans ce chapitre sera comparée, dans le premier manuscrit (chapitre 4), à différents logiciels commerciaux d'analyse de gels DGGE. Les résultats produits par cette approche seront par la suite utilisés par le modèle d'élongation présenté également dans le premier manuscrit (chapitre 4) et validé dans le second manuscrit (chapitre 5).

3.1 Délimitation et soustraction du bruit de fond

L'analyse des profils DGGE débute par la délimitation et la soustraction du bruit de fond de l'échantillon. Tel que présenté à la Figure 3-1, les profils DGGE correspondant à des sols non contaminés sont complexes, présentent une forte superposition des bandes ainsi qu'un bruit de fond important. La nature imparfaite des profils DGGE expérimentaux complique également la situation.

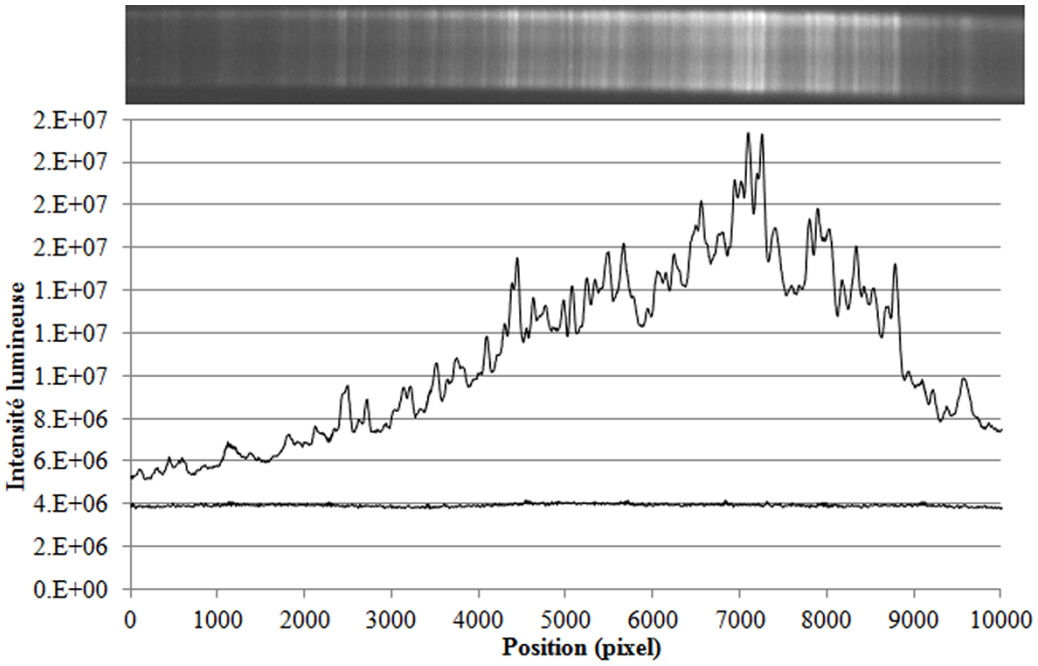

Figure 3-1 : Profil DGGE et courbe d'intensité lumineuse (en niveaux de gris) associée. Sur le graphique, la courbe visible à une intensité lumineuse de ~$4*10^6$ correspond à un profil DGGE où un échantillon PCR ne contenant pas d'ADN a été injecté. Afin de faciliter l'analyse, les images de gels DGGE sont redimensionnées en utilisant un facteur de 10 (10 240 pixels au lieu de 1024). Il a effectivement été observé ce redimensionnement facilitait la convergence du script d'optimisation décrit à la section 3.2.

La première étape de l'analyse consiste à soustraire la portion de signal correspondant au bruit de fond. Puisque l'intensité de base liée au profil des échantillons à analyser est plus élevée que celle générée par le blanc PCR (Figure 3-1), le profil de bruit de fond des échantillons a été délimité en reliant l'intensité du signal observé aux deux extrémités des profils. Ce bruit de fond, généré par l'interaction du gel avec la molécule utilisée afin de colorer ce dernier ou par le passage dans le gel du produit amplifié par PCR (qui peut contenir des nucléotides ainsi qu'une certaine quantité de courts fragments d'ADN ou de dimères d'amorces, etc.), ne sera pas considéré dans les analyses subséquentes.

Une fois que le bruit de fond est soustrait, il faut délimiter la portion de signal attribuable aux espèces abondantes (pics visibles) de celle générée par la multitude d'espèces n'étant pas suffisamment abondantes pour produire une bande visible sur le gel. Contrairement aux approches généralement employées afin d'analyser les profils DGGE, l'approche présentée dans ce chapitre

ne considère pas le signal situé sous les pics visibles comme du bruit de fond. Cette portion de signal est plutôt associée à toutes les espèces ciblées par les amorces DGGE mais trop peu abondantes pour produire une bande visible sur le gel. Il a été montré que le ratio « intensité sous les pics »/« intensité sous le profil en entier », appelé ci-après *peak-to-signal ratio* (PSR), était relié à la richesse des communautés étudiées (Loisel et al., 2006). Afin de produire des résultats quantitatifs, il est donc primordial d'être en mesure de délimiter le signal attribuable aux pics visibles et le reste du profil le plus précisément possible.

Plusieurs approches automatisées, basées sur des algorithmes d'optimisation considérant la forme des bandes DGGE, ont été développées et testées mais aucune d'entre-elles ne fut en mesure de produire des résultats satisfaisants pour des profils DGGE expérimentaux. Dès lors, il fut déterminé que la meilleure, et possiblement la seule, approche permettant de déterminer à quel niveau d'intensité se situent la base des bandes visible est un ajustement manuel basé sur l'observation attentive du profil. L'analyse des profils DGGE *in silico* a permis d'établir que la racine des bandes visible se situait immanquablement très près des minimums visibles sur le profil. L'ajustement manuel doit donc majoritairement se faire dans les zones où les bandes sont superposées.

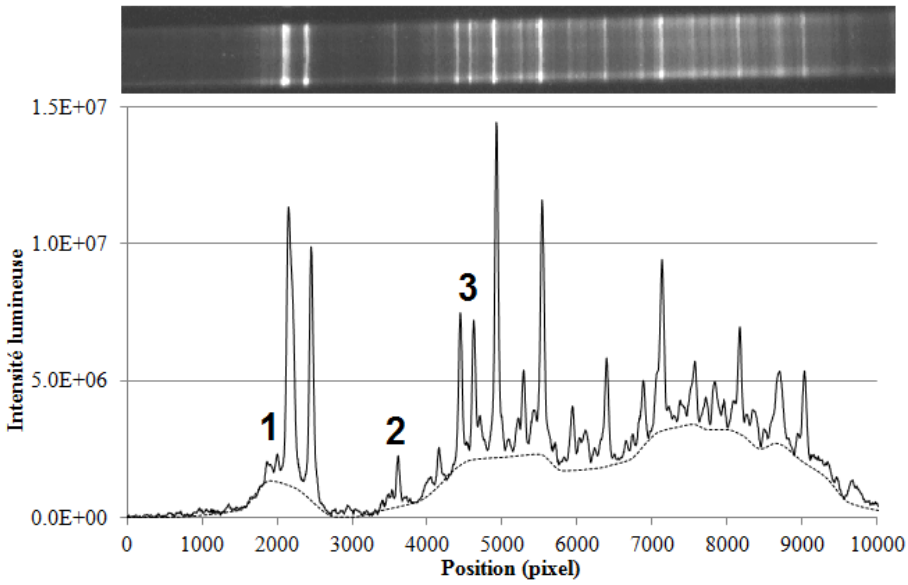

Figure 3-2 : Délimitation entre le signal attribuable aux espèces abondantes (pics visibles) et celui attribuable aux espèces plus rares. La signification des chiffres 1, 2 et 3 est expliquée dans le texte plus bas.

D'entrée de jeu, il est important de mentionner que le profil permettant de délimiter les pics dépend de la composition des séquences d'ADN d'un échantillon donné. Dès lors, ce profil peut présenter de brusques variations à certaines positions sur le gel. L'ajustement du profil doit se faire de façon à ce que l'importance relative de chaque bande visible soit logique. Par exemple, sur la Figure 3-2, les pics identifiés à l'aide du chiffre 1 correspondent à des bandes pâles, floues et mal définies. Au contraire, le pic identifié par le chiffre 2, sans être très brillant, est très bien défini. Le profil de délimitation des pics doit donc être ajusté afin que l'importance des pics situés en 1, influencés à la hausse par la présence d'un pic dominant, soit inférieure à l'importance du pic situé en 2. De la même façon, à la position 3 se trouvent une série de quatre pics parmi lesquels uniquement deux sont brillants. Si le profil de délimitation des pics était situé trop bas, l'importance des pics très pâles serait disproportionnée par rapport à celle des pics abondants. Le profil de délimitation est donc ajusté de façon à ce que l'importance d'un pic ne soit pas influencée à la hausse ou à la baisse par sa position sur le gel.

3.2 Quantification des pics

L'algorithme de quantification des pics DGGE est basé sur l'observation que les pics DGGE correspondant à des espèces uniques étaient très bien représentés par des distributions de densité de probabilités Gaussiennes (Eq.1).

$$I(x) = \frac{A}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_0)^2}{2\sigma^2}} \qquad [1]$$

Les pics DGGE sont donc complètement représentés à l'aide de trois paramètres; la position centrale du pic sur le gel (x_0 en pixels), son amplitude (A, en niveaux de gris) ainsi que sa largeur représentée par l'écart-type de la distribution (σ en pixels). Un algorithme d'optimisation a d'abord été développé à l'aide du logiciel Matlab afin de déterminer automatiquement l'amplitude et l'écart-type les plus représentatifs d'un pic unique donné. Cet algorithme nécessite comme paramètres d'entré un estimé initial et des bornes minimales et maximales pour l'amplitude et l'écart-type du pic, ainsi qu'un estimé de la position centrale du pic (déterminé visuellement sur le gel). L'algorithme, basé sur la fonction *fmincon* de Matlab, optimisera ensuite ces paramètres de façon à représenter le pic le plus fidèlement possible.

Lorsque les résultats furent satisfaisants pour la quantification d'un pic unique, l'algorithme fut généralisé de façon à pouvoir optimiser plusieurs pics simultanément. L'analyse des profils DGGE est interactive (Figure 3-3). Dans une boîte de dialogue apparaissant à l'écran, l'utilisateur fournit un estimé initial de la position centrale de tous les pics ainsi qu'une délimitation des pixels devant servir au processus d'optimisation (début et fin). Afin d'assurer la convergence de l'algorithme, moins de 10 pics peuvent être quantifiés simultanément. L'algorithme optimise l'amplitude et l'écart-type correspondant à chaque pic, tout en permettant aux positions centrales visuellement estimées de varier de quelques pixels vers la gauche ou vers la droite. Les résultats sont présentés sur un graphique et l'analyste peut accepter ou refuser ces derniers. S'il y a refus, de nouvelles positions centrales peuvent être fournies et/ou plus ou moins de pics peuvent être considérés simultanément. L'analyse se poursuit jusqu'à ce que tous les pics aient été quantifiés. L'abondance relative des différentes bandes est ensuite déterminée à partir des amplitudes optimisées. Les résultats générés par l'analyse des profils DGGE (bruit de fond et abondance relative des bandes) seront utilisés par le modèle d'élongation des RAD DGGE présenté au chapitre suivant.

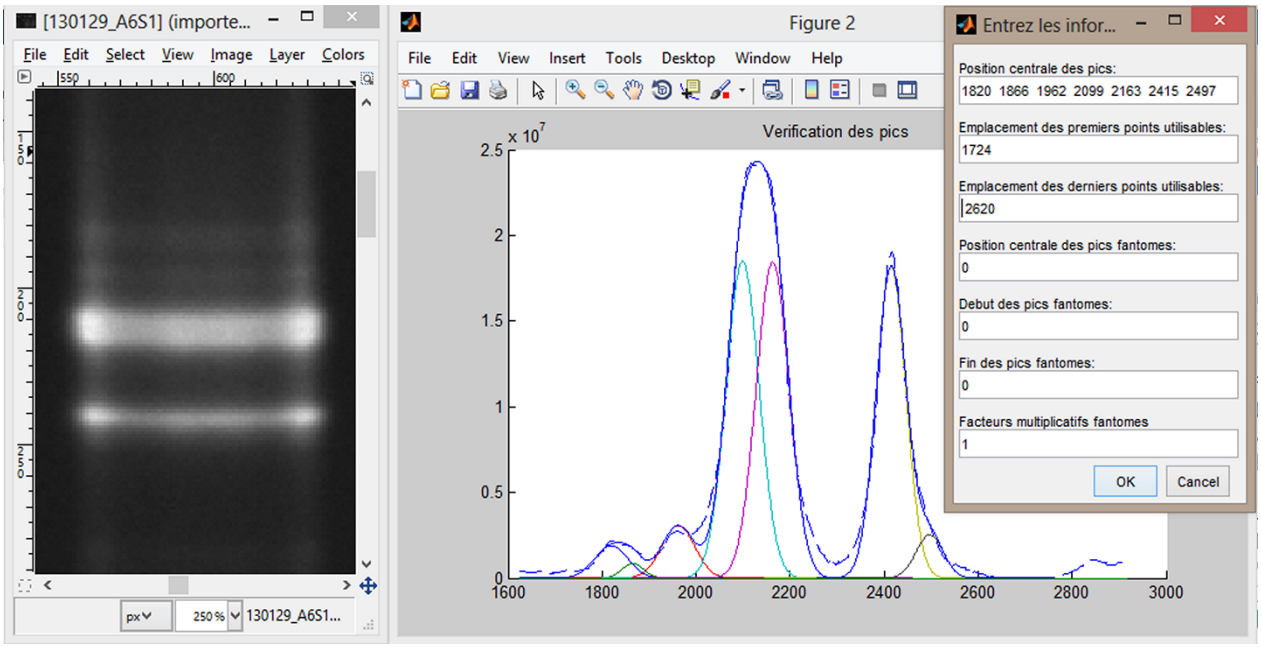

Figure 3-3 : Résultats produits par l'algorithme de quantification des pics DGGE. À la gauche de la figure se trouve la portion de profil à analyser. L'information par rapport à la position des pics est déterminée directement à partir du profil analysé et est entrée dans une boîte de dialogue (à droite de la figure). L'algorithme optimise alors les paramètres des distributions correspondant à chaque pic et présente les résultats sur un graphique. Les résultats peuvent être acceptés ou rejetés.

CHAPITRE 4 DÉVELOPPEMENT D'UNE MÉTHODOLOGIE PERMETTANT DE QUANTIFIER LA DIVERSITÉ DES COMMUNAUTÉS BACTÉRIENNES DES SOLS À PARTIR DU DGGE

Cette section est centrée autour du premier manuscrit intitulé « *A new framework to accurately quantify soil bacterial community diversity from DGGE* » (section 4.2) publié dans la revue *Microbial Ecology* (soumis le 19 décembre 2012, accepté le 11 avril 2013).

4.1 Présentation du premier manuscrit

Plusieurs logiciels commerciaux sont disponibles afin d'analyser les gels DGGE. En communiquant avec les développeurs de ces logiciels, il est possible de réaliser que ces derniers n'ont pas été développés spécifiquement afin d'analyser des profils complexes. L'approche analytique présentée au chapitre 3 a été développé dans le but de contourner ces difficultés et permettre l'analyse de profils DGGE complexes. Afin de prouver son utilité, ce manuscrit vise à comparer ses résultats à ceux produits par trois logiciels commerciaux différents. Le manuscrit cherche également à apporter des réponses à plusieurs questions concernant l'emploi du DGGE à des fins quantitatives. Tel que présenté à la section 1.4.2.1 (**La phylogénie microbienne à l'ère de la microbiologie moléculaire**), la notion d'espèce en microbiologie est intimement liée à la similitude qui existe entre l'ADN des différents organismes. Le processus de formation des profils DGGE est différent. En théorie, deux séquences qui diffèrent d'une seule base migreront à des positions différentes. En pratique, lorsque la communauté étudiée est composée de milliers d'espèces, la position de migration de plusieurs espèces peut être identique. La similitude existant entre le processus de formation des profils DGGE et le processus de groupement des séquences d'ADN en OTU n'a jamais été étudiée.

Réalisant que l'approche analytique développée dans le cadre de ce projet permettait de produire des résultats similaires à ceux produits en regroupant en OTU les séquences d'ADN similaires à 98% ou plus, une méthode d'élongation a été développée. Cette méthode permet, à partir des RAD partielles issues de la quantification des pics DGGE et des PSR extraits des profils, d'estimer la diversité des communautés bactériennes étudiées.

4.2 "A new framework to accurately quantify soil bacterial community diversity from DGGE"

4.2.1 Abstract

Denaturing gradient gel electrophoresis (DGGE) has been and remains extensively used to assess and monitor the effects of various treatments on soil bacterial communities. Considering only abundant phylotypes, the diversity estimates produced by this technique have been proven to be uncorrelated to true community diversity. The aim of this paper was to develop a framework to estimate a community's true diversity from DGGE. Developed using *in silico* DGGE profiles generated from published pyrosequencing datasets, this framework elongates the rank-abundance distributions (RADs) drawn by band quantification using the peak-to-signal ratio (PSR) parameter, which was proven to be related to bacterial richness. The ability to compare DGGE-based diversity estimates to the true diversity of communities led to a unique opportunity to identify potential pitfalls when analyzing DGGE gels with commercial analysis software programs and gain insight into the process of DNA band clustering in the profiles. Bacterial diversity was compared through richness, Shannon and Simpson's 1/D indices. Intermediate results demonstrated that, even though commercial gel analysis software programs were unable to produce consistent results throughout all samples, a newly developed Matlab-based framework unravelled the dominance profiles of communities from band quantification. Elongating these partial RADs using the PSRs extracted from the DGGE profiles chiefly made it possible to accurately estimate the true diversity of communities. For all the samples analyzed, the estimated Shannon and Simpson's 1/D were accurate at ±10%. Richness estimations were less accurate, ranging from -11 to 31% of the expected values. The framework showed great potential to study the structure and diversity of soil bacterial communities.

4.2.2 Introduction

Considering the importance of diversity with regards to ecosystem functioning (Hooper et al., 2005) and its resistance and resilience to perturbations (Girvan et al., 2005), it is of great importance to be able to routinely assess and compare soil microbial community diversity across a large scale of management practices and treatments.

Indeed, soil microbial communities are among the most diverse and abundant on Earth (Roesch et

al., 2007) and play a key role in terrestrial ecosystems (van der Heijden et al., 2008). Even if these communities can be studied in great depth using modern metagenomics (Simon and Daniel, 2011), researchers still require cost-effective methods to reliably estimate the diversity of multiple samples. Fingerprinting techniques such as denaturing gradient gel electrophoresis (DGGE) have been successfully used in many diversity studies (Nakatsu, 2007).

For such complex communities, interpreting the outcome of DGGE-based diversity surveys—even though they are of widespread use—is not an easy task. Initially developed as a mutation detection tool, DGGE can theoretically separate DNA sequences that differ by a single base pair (Myers et al., 1987; Sheffield et al., 1989). Consequently, in estimating community diversity from DGGE profiles, visible bands are implicitly associated with unique phylotypes. Since some authors (Schmalenberger and Tebbe, 2003) have clearly demonstrated that DGGE bands may contain many different operational taxonomic units (OTUs), this general consideration is known to be false. Generally speaking, the similarity level between the genetic markers of different organisms constitutes the baseline of OTU definition in modern microbial taxonomy. It has been stated that the minimal value of 97% similarity between 16S rRNA gene sequences (Stackebrandt and Goebel, 1994) may lack the resolving power to relate bacteria at the species level (Rosselló-Mora and Amann, 2001), especially for short partial sequences. If the similarity level that should be used is debatable, the concept still provides a useful and scientifically grounded basis to estimate and compare bacterial community diversity. Even though DNA band superposition on DGGE profiles has been shown to occur, it is not known whether this process can be related to DNA sequence clustering at any similarity level or if both processes will consistently yield a comparable representation of soil bacterial community dominance profiles.

Another poorly understood and seemingly overlooked aspect of DGGE is the profile analysis step itself. In the scientific literature, fingerprinting patterns were analyzed using many different software programs. Although convenient, it must be recognized that most were not specifically developed to analyze fingerprints as complex as those produced by soil bacterial communities (personal communications). It seems that the differences between software programs and more generally their capacity to quantitatively unravel the communities' true dominance profile from banding patterns have never been assessed.

Furthermore, numerical simulations demonstrated that by only enabling the consideration of the

most abundant phylotypes, fingerprinting techniques provide an inaccurate estimation of the true diversity of microbial communities (Blackwood et al., 2007; Narang and Dunbar, 2004). The diversity indices routinely used to quantify diversity are influenced by both the richness (length) and the dominance pattern conveyed by the dataset used to plot the rank-abundance distribution (RAD) of the studied community (Hill et al., 2003; Magurran, 2004). Traditional DGGE banding pattern analysis theoretically yields information on community dominance profiles but not on richness. However, based on simulated bacterial communities, it was shown that the subunit background percentage (SBP), an indicator extractable from DGGE profiles, was related to community richness (Loisel et al., 2006). The SBP can be seen as a measure of the proportion of the studied community that is not accounted for when only visible peaks are considered. Unfortunately, the relationship between the SBP and richness is not straightforward and was shown to depend on the abundance model to generate the simulated bacterial communities. Used in conjunction with theoretical abundance models, the SBP could make it possible to infer whole community RADs from DGGE banding patterns. The most commonly used models to describe soil bacterial communities are the lognormal (Blackwood et al., 2007; Doroghazi and Buckley, 2008; Dunbar, Barns, Ticknor, and Kuske, 2002; Loisel et al., 2006; Narang and Dunbar, 2004), the power law (Gans, Wolinsky, and Dunbar, 2005; İnceoğlu, Al-Soud, Salles, Semenov, and van Elsas, 2011) and the geometric (Loisel et al., 2006; Narang and Dunbar, 2004) distributions, and these abundance models could be used to elongate the partial DGGE-based RADs to "add" the OTUs that are not abundant enough to produce a visible band on the gel. The SBP would provide a stopping criterion for the elongation process, indicating the length at which the distribution is complete. If such a methodology could be developed, it would allow for the accurate characterization of soil bacterial community diversity from DGGE fingerprints.

The main objective of this paper was to develop such a framework. In the process of developing this framework, the influence of certain analytical parameters (background noise subtraction and DGGE peaks quantification) and the extent of the bands clustering process were assessed. To do so, publicly available pyrosequencing datasets of soil bacterial communities (Nacke et al., 2011) were used to generate *in silico* DGGE profiles. Knowing community composition made it possible to assess whether DGGE-based diversity estimations can theoretically lead to similar conclusions than more robust methods based on DNA sequencing and clustering.

4.2.3 Methods

4.2.3.1 In silico DGGE profiles construction

In silico DGGE profiles were constructed using pyrosequencing datasets of soil bacterial communities downloaded from the NCBI Sequence Read Archive (NCBI, 2012). This methodology was selected to avoid the necessity of choosing a suitable theoretical model to derive the RADs used to create the profiles. Datasets were generated using DNA extracted from six different soils and targeting the V2-V3 region of the bacterial 16S rRNA gene (approximately 400 nt) (Nacke et al., 2011). The sequences were processed by the authors to remove the primers and the regions presenting low quality scores and ranged in length between 200 and 300 nt.

Presenting different initial richness and dominance patterns, three of these datasets were selected for further analysis: FUG3 (intensely fertilized grassland), BF2 (unmanaged beech forest) and SAF1 (spruce forest). Each dataset was aligned and clustered using three similarity levels (100, 97 and 95%) with the RDP pyrosequencing pipeline (Cole et al., 2009). Datasets were further simplified by associating the relative abundance of every OTU with a unique representative sequence, ensuring that every cluster generates a unique DGGE band in the *in silico* profiles. A total of nine communities were produced, ranging in richness between 1895 and 17552 OTUs and presenting a most dominant OTU relative abundance between 1.5 and 5.4%.

OTUs were positioned in the profiles using DNA sequence theoretical melting temperatures (T_m) calculated with a published predictive model (Khandelwal and Bhyravabhotla, 2010). This model was selected because it gave good results over a wide range of sequence lengths (15-mers to genomic). Gel conditions (denaturing gradient) were adjusted considering that urea and formamide reduce DNA sequence melting temperatures by 2.25°C/M and 0.6°C/% (Hutton, 1977), respectively, ensuring that all the sequences were included in the profiles. DGGE peaks were represented in the *in silico* profiles by Gaussian probability density functions (PDFs) (Eq.1).

$$I(x) = \frac{A}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_0)^2}{2\sigma^2}} \quad [1]$$

The peak corresponding to each OTU was therefore completely represented by three parameters: central position on the gel (x_0 in pixels, determined from the T_m), amplitude (A in greyscale intensity proportional to the OTU relative abundance in its dataset RAD) and peak width (σ in pixels,

standard deviation). Based on observations of experimental DGGE gels, the standard deviation was set at a mean value of 2.0 pixels and forced to vary randomly for every OTU between ±10% of the mean value. Peak intensity *I(x)* was evaluated for all of the pixels (*x* values) over the whole gel vertical length (set at 1024 pixels). The *in silico* profiles were obtained by summing up the intensity corresponding to every OTU contained within a given dataset for every pixel.

The image representing the *in silico* DGGE gel was created in 16 bits uncompressed TIFF format (maximum greyscale value of 65535). The intensities of all the vertical profiles were first normalized so that the maximum greyscale value for every sample equals 50000. This step was seen as analogous to adjusting the exposition time when photographing DGGE gels. Two-dimensional profiles were considered to be 175 pixels wide, all filled with the previously generated one-dimensional profiles. In order to reproduce some of the difficulties associated with the analysis of real DGGE gels, an additional background noise calculated from the profiles mean intensity level and randomly adjusted was added.

4.2.3.2 Profiles analysis

In silico DGGE profiles were analyzed using four different software programs: TotalLab Quant (TotalLab Ltd., Newcastle upon Tyne, UK), GelCompar II (Applied Maths, Inc., Austin, USA), BIO-1D advanced (Vilber Lourmat, Marne-la-Vallée, France) and a Matlab-based program (The MathWorks Inc., Natick, USA) specifically developed for this paper. The main differences between software programs were mostly associated to their background subtraction or peak delimitation algorithms. These parameters were therefore further studied.

Background subtraction

The background subtraction algorithms evaluated in this paper were limited to the popular rolling ball approaches included in TotalLab, GelCompar II and BIO-1D and the approach developed for the Matlab-based program. TotalLab Quant and GelCompar II include practically the same algorithm: a virtual ball whose diameter is chosen by the user rolls under the profiles and subtracts the signal located under the top of the ball. BIO-1D's algorithm is slightly different, since it first subtracts the signal from the center of the ball and then asks the user to define a threshold level to adjust profile baseline intensities to zero.

GelCompar II and BIO-1D both propose an "optimal" ball size for a given gel (41 and 72 pixels

for the synthesized image, respectively). Every profile was therefore analyzed with each software program using ball radiuses of 20, 41, 72 and 144 (94 for BIO-1D since it was the maximum size allowed by the software). TotalLab Quant and BIO-1D do not require the redefinition of the peaks when modifying the ball size, and additional radiuses of 5 and 10 were added for these software programs.

Background subtraction with the Matlab-based program worked differently. After trying to develop an automated calculation procedure to derive background noise profiles from DGGE gels, it was observed that a manual adjustment was the best and, perhaps, the only way to correctly draw a line between peaks area and background. This manual adjustment was based on the careful observation of the image of the gel. The background profiles were derived by qualitatively ranking neighboring peaks from very weak to very bright. It was observed that a background profile very close to the peaks' root gives more weight to the brightest peaks and vice versa. Adjusting the background level closer to or farther from the peaks' root made it possible to draw the most representative picture of what is visually conveyed by the image. It must be mentioned that this process is iterative. During the quantification process, if a peak is disproportionate as compared to its neighbors or if its optimized standard deviation is significantly different from all the other peaks, it may be necessary to adjust the background profile accordingly.

Peaks delimitation and quantification

In the four considered software programs, peaks were quantified using two different general approaches. The approach shared by TotalLab Quant and BIO-1D consists in delimiting peaks with two straight lines. Peaks are then quantified by summing up the intensity of the background-subtracted profile between these lines. In contrast, GelCompar II and the Matlab-based framework adjust Gaussian PDFs under the peaks (Eq.1). This adjustment is done manually in GelCompar II, while the Matlab-based framework automatically and simultaneously optimizes many peaks. For the Matlab-based framework, profile analyses are conducted in many optimization rounds. In an interactive dialog box, the analyst enters information on the central position of the peaks to be quantified (fewer than 10 peaks for every round). Central positions are determined directly from the image of the DGGE gel viewed with any image editing software program. After the convergence of the algorithm, optimized peaks are plotted against the analyzed profile. The analyst can accept or reject the results. If rejected, the optimization routine can be run again with different

initial central positions or with fewer or more peaks. The routine is run repeatedly until the peaks are accepted. The resulting PDF parameters are then saved and the optimization moves on to other peaks, until the entire profile is analyzed. Peak abundances are finally determined from the PDF amplitudes.

4.2.3.3 Representativeness of DGGE-based dominance profiles

While a similarity level of 97% is chosen in almost all sequencing-based bacterial community diversity surveys, it is doubtful that the value has any meaning when analyzing DGGE banding patterns. If all the DGGE bands were generated by a single OTU, RADs drawn from peak quantification would be very similar to RADs produced by sequencing dataset clustered at the 100% similarity level. Considering that DGGE bands are known to superpose to a certain extent, datasets used to generate *in silico* profiles were further clustered using the RDP pipeline (Cole et al., 2009), with similarity levels ranging from 96 to 100%. This step aims to determine whether the DNA band superposition process is numerically similar to the clustering of DNA sequences at a specific similarity level. For simplicity, in the particular context of this publication, RADs and diversity indices calculated from the pyrosequencing datasets used to generate the *in silico* DGGE profiles will be referred to as the *true RADs* and *true diversity* for a certain similarity level.

The peak-to-signal ratio (PSR=1-SBP), a parameter analogous to the SBP (Loisel et al., 2006), was extracted from all the *in silico* DGGE profiles. For each sample, the PSR was calculated as the area under all the peaks divided by the area under the whole profile. The background noise added under the DGGE profiles when synthesizing the image was subtracted before calculating the PSRs. This parameter represents the percentage of all the DNA sequences loaded into a DGGE profile contained within the most abundant OTUs (the peaks). The remainder belongs to the OTUs not abundant enough to produce a visible band on the gel. These OTUs are incidentally unaccounted for in the diversity estimates produced through DGGE and will be dealt with by the elongation framework (introduced below).

Finally, considering that small peaks may not be very indicative of true community dominance profiles, DGGE-based RADs were truncated by subtracting all the peaks with relative abundances smaller than a certain cut-off. Since it was impossible to objectively choose an appropriate cut-off, percentages between 0 and 3.0% (in 0.2% increments) were successively used in order to identify the optimal value for subsequent analyses. This truncation was judged necessary because it was

observed that DGGE-based and true RADs were deviating at a certain relative abundance value. The PSR values were modified to take into account the peaks that were removed from the DGGE-based RADs.

To compare DGGE-based and true community dominance profiles, true RADs had to be modified. For all the samples, the true distributions were truncated by keeping the same number of OTUs as the number of peaks above the cut-off percentage. True PSRs were then calculated as the number of sequences in the truncated RADs divided by the number of sequences in the complete distributions. Finally, since DGGE band quantification yields results in relative abundance, true RADs were transformed accordingly. This truncation lead to the calculation of biased diversity indices and was solely used for comparative purposes. The representativeness of DGGE-based dominance profiles was evaluated using four indicators. The first indicator was used to verify whether PSRs can be accurately extracted from DGGE profiles:

1. ΔPSR: Deviation percentage of DGGE-based PSRs compared to true PSRs calculated from sequencing dataset clustering;

Since the main objective of this paper was to determine whether DGGE could be confidently used to assess the diversity of soil bacterial communities, two ubiquitous diversity indices were calculated to characterize the RADs representing the communities' dominance profiles. These indices were calculated using PAST software (Hammer, Ryan, and Harper, 2001).

2. ΔH': Deviation percentage of DGGE-based Shannon indices from corresponding expected values;
3. Δ1/D: Deviation percentage of DGGE-based Simpson's 1/D indices from corresponding expected values.

In order to further characterize the similarity between DGGE and clustering-based dominance profiles, the Euclidean distance (Legendre and Legendre, 1998) was calculated. Unlike the two previous diversity indices, this measure associated every DGGE peak to its principal underlying OTU. To do so, theoretical OTU migratory positions were associated with actual peak locations in the *in silico* profiles. Since the measure was calculated using the truncated RADs produced by both approaches, certain peaks did not match any OTU theoretical position and *vice-versa*. In these cases, the relative abundance of the corresponding peak/OTU was set at 0%.

4. $D_{EUCLIDEAN}$: Calculated using Eq.2, where A_{DGGE} corresponds to peak relative intensities, A_{OTU} represents OTU relative abundances and n corresponds to RAD lengths.

$$D_{EUCLIDEAN} = \sqrt{\sum_{i=1}^{n}(A_{iDGGE} - A_{iOTU})^2} \quad [2]$$

4.2.3.4 Using PSRs to improve DGGE-based diversity estimates

Using PSR values extracted from the profiles, an empirical framework to estimate true community diversity from DGGE was developed. Since DGGE and pyrosequencing-based dominance profiles were very similar when clustering the datasets at the 98% level, the framework was developed to estimate true community diversity at that particular similarity level.

First of all, RADs produced by the Matlab-based framework were truncated using the optimal cut-off value (1.0%). After truncation, RADs were normalized (sum = 1), multiplied by corresponding PSRs and further multiplied by 35000. This last step aimed to make it possible to work in terms of absolute rather than relative abundance. The value 35000 was chosen because it was close to the number of reads per sample in the pyrosequencing datasets used in this publication.

Distributions were then elongated to "add" the species that were not accounted for in the peak quantification process. This elongation framework was designed and calibrated using only the true community RADs. The method was then applied without further modification over the DGGE results. Although the lognormal, power law and geometric distributions are the most commonly used abundance models to describe soil bacterial communities, they were unable to fit the true RADs correctly. The power law distribution (PLD, Eq.3) provided an acceptable fit if a distinct model parameterization is used to predict mid and low abundance values. Starting right after the last DGGE peak above the optimal cut-off value, the elongation framework was therefore divided into two distinct steps, both based on Eq.3.

$$PLD(x) = x_{min} * x^{-\alpha} \quad [3]$$

The first step elongated the truncated DGGE-based RADs until a richness of 699 was reached using a PLD exponent (α) of 0.875 for all the samples. The abscissa (x value in Eq.3) producing an abundance just below that of the last retained peak was selected as the starting x value for the elongation. This initial abscissa was distinct for all samples and once determined was increased by one each time a species was added to the RAD. The other parameter of Eq.3, x_{min}, varied between

500 and 3000 and was optimized for each sample to ensure continuity in the predicted abundance values at the junction of the two elongation steps. The second elongation step produced abundance values for species of rank 700 and more. In this second step, species rank corresponded to abscissa values. The PLD parameters α and x_{min} are functions of the PSR values and were determined using relationships derived from the true community RADs (Eq. 4 and 5). For both elongation steps, predicted abundance values were rounded to the nearest integer. Considering that the elongation framework was designed to work in absolute abundance, when a value of 1 was reached, singletons were added until the sum of the abundance of all the species equalled 35 000. Figure 4-1 presents a schematic representation of the elongation framework.

$$\alpha = 0.267 * PSR - 0.935 \quad [4]$$

$$x_{min} = -5034 * PSR + 3656 \quad [5]$$

It must be emphasized that the elongation framework produces distributions that do not follow a particular abundance model. Indeed, RAD heads are drawn from peak quantification and are distribution free. Mid- and low-abundance values are all predicted using the PLD, but with different model parameterization. Therefore, resulting RADs do not generally follow a power law.

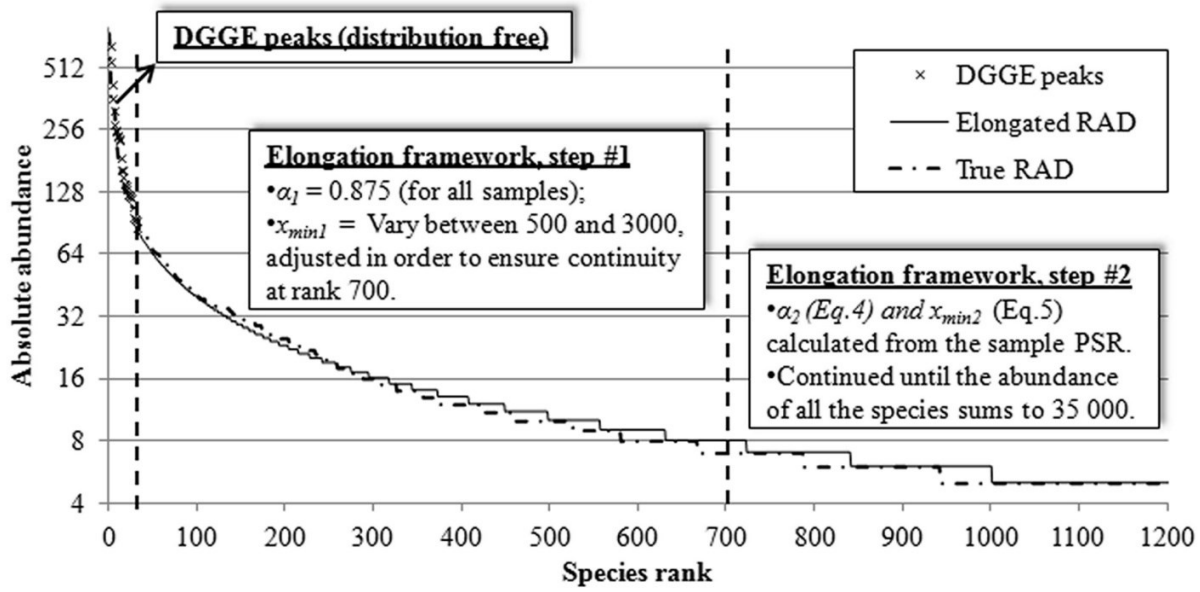

Figure 4-1 : Schematic representation of the elongation framework. The framework was developed with pyrosequencing datasets containing approximately 35 000 sequences per sample. The elongated RADs were therefore considered complete when the abundance of all the species summed to this value. The example is given for sample $BF_{100\%}$, and the results of the elongation process may be compared with the true RAD. Results are only presented for species rank under 1200 for readability but the actual richness of this sample was around 9500.

Using these elongated RADs, bacterial community diversity was characterized through three indicators: community richness (number of species), Shannon and Simpson's 1/D indices. Estimated diversities were compared to true community diversity at the 98% similarity level (untruncated RADs). All indices were calculated using PAST software (Hammer et al., 2001).

4.2.4 Results

4.2.4.1 Representativeness of DGGE-based dominance profiles

Pyrosequencing datasets of 16S rRNA gene sequences from three different environments were clustered at 95%, 97% and 100% similarity level, generating nine "theoretical" bacterial communities. Based on the theoretical T_m of a unique representative sequence selected for each OTU, *in silico* DGGE profiles were derived (Figure 4-2).

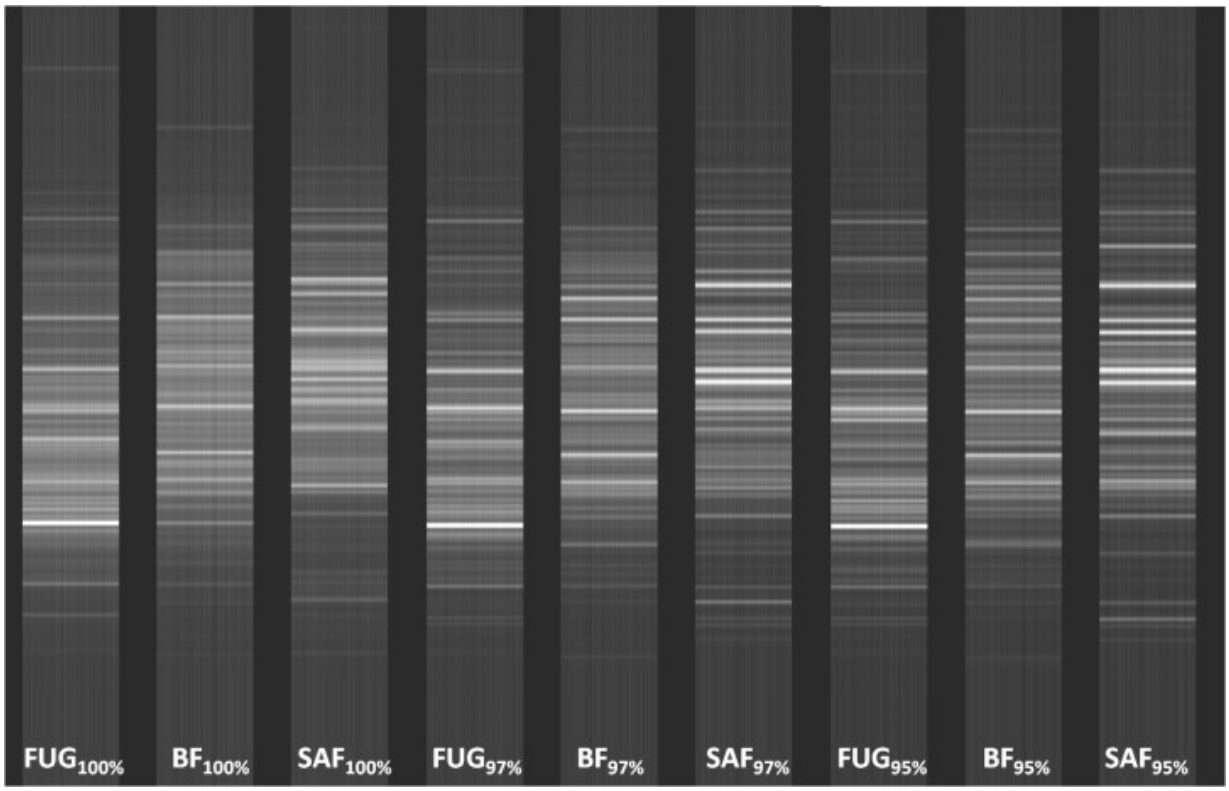

Figure 4-2 : Image of the gel synthesized from the datasets published by Nacke et al. (2011).

Each *in silico* profile was analyzed in 17 ways (software programs and ball sizes), and resulting RADs were truncated using various cut-off values and compared to corresponding pyrosequencing datasets clustered using 5 different similarity levels (from 96 to 100%). This methodology generated a significant amount of data, and the complete results are presented as supplementary material (TableauA3-1.xlsx). To synthesize the results, the parameters that were found to be optimal for all the sample-software pairs are presented in Tableau 4-1. Optimal parameters correspond to the ball size and similarity level that made it possible to meet predefined criteria (ΔPSR, ΔH' and $\Delta 1/D \leq \pm 10\%$) over the widest range of cut-off values. Indicator stability for different cut-off values was deemed to be a very important aspect to consider, since chance alone can yield good results for a specific cut-off percentage. A more complete version of Tableau 4-1, which also presents the mean values of the indicators over the entire reported cut-off range, is presented as supplementary material (Tableau A3-2).

Tableau 4-1 : For the Matlab-based framework, TotalLab Quant, GelCompar II and BIO-1D, optimal parameters obtained for all the samples. Ball size and similarity level values were selected as those simultaneously generating ΔPSR, ΔH' and Δ1/D values lower than ±10% over the widest range of cut-off values.

Software program	Sample	Optimal ball size	Closest similarity level	Cut-off range
Matlab-based framework	$FUG_{100\%}$	-	98%	0.4 - 3.0
	$BF_{100\%}$	-	98%	0.4- 3.0
	$SAF_{100\%}$	-	98%	0.4 - 1.8
	$FUG_{97\%}$	-	98%	0.6 - 3.0
	$BF_{97\%}$	-	97%	1.2 - 3.0
	$SAF_{97\%}$	-	98%	0.6 - 3.0
	$FUG_{95\%}$	-	98%	0.4 - 3.0
	$BF_{95\%}$	-	98%	0.4 - 3.0
	$SAF_{95\%}$	-	98%	0.2 - 3.0
TotalLab Quant	$FUG_{100\%}$	10	96%	2.2 - 3.0
	$BF_{100\%}$	5	100%	0.8 - 3.0
	$SAF_{100\%}$	10	98%	0.8 - 2.2
	$FUG_{97\%}$	5	100%	2.8 - 3.0
	$BF_{97\%}$	5	96%	2.0 - 3.0
	$SAF_{97\%}$	5	97%	2.8 - 3.0
	$FUG_{95\%}$	None	-	-
	$BF_{95\%}$	10	98%	1.0 - 3.0
	$SAF_{95\%}$	10	97%	1.0 - 3.0

Tableau 4-1 : For the Matlab-based framework, TotalLab Quant, GelCompar II and BIO-1D, optimal parameters obtained for all the samples. Ball size and similarity level values were selected as those simultaneously generating ΔPSR, ΔH' and Δ1/D values lower than ±10% over the widest range of cut-off values (continued).

Software program	Sample	Optimal ball size	Closest similarity level	Cut-off range
GelCompar II	$FUG_{100\%}$	41	96%	0.6 - 1.8
	$BF_{100\%}$	72	96%	1.0 - 3.0
	$SAF_{100\%}$	72	97%	0.4 - 1.6
	$FUG_{97\%}$	20	96%	0.6 - 3.0
	$BF_{97\%}$	20	96%	0.6 - 0.8
	$SAF_{97\%}$	72	96%	1.0 - 3.0
	$FUG_{95\%}$	20	96%	0.2 - 3.0
	$BF_{95\%}$	20	100%	0.4 - 3.0
	$SAF_{95\%}$	20	98%	0.6 - 3.0
BIO-1D	$FUG_{100\%}$	None	-	-
	$BF_{100\%}$	5	99%	0.4 - 3.0
	$SAF_{100\%}$	10	100%	0.8 - 3.0
	$FUG_{97\%}$	None	-	-
	$BF_{97\%}$	41	97%	0.0 - 3.0
	$SAF_{97\%}$	41	99%	0.8 - 3.0
	$FUG_{95\%}$	20	97%	0.0 - 0.2
	$BF_{95\%}$	None	-	-
	$SAF_{95\%}$	41	100%	0.6 - 3.0

Tableau 4-2 also presents the parameters that were deemed optimal for each software program. However, since DGGE profile analysis traditionally involves the use of a single ball size for all the samples loaded on a given gel, the parameters presented in Tableau 4-2 were selected considering all nine samples simultaneously. The optimal cut-off percentage was also limited to a single value. Figure 4-3 presents the mean ΔPSR, ΔH', Δ1/D and $D_{EUCLIDEAN}$ values produced by each software program when the parameters presented in Tableau 4-2 are used.

Tableau 4-2 : Parameters that made it possible to minimize indicator values when simultaneously considering all the samples.

Software program	Optimal ball size	Optimal cut-off value	Closest similarity level
Matlab-based framework	-	1.0%	98%
TotalLab Quant	20	1.6%	96%
GelCompar II	41	1.0%	96%
BIO-1D	41	0.4%	98%

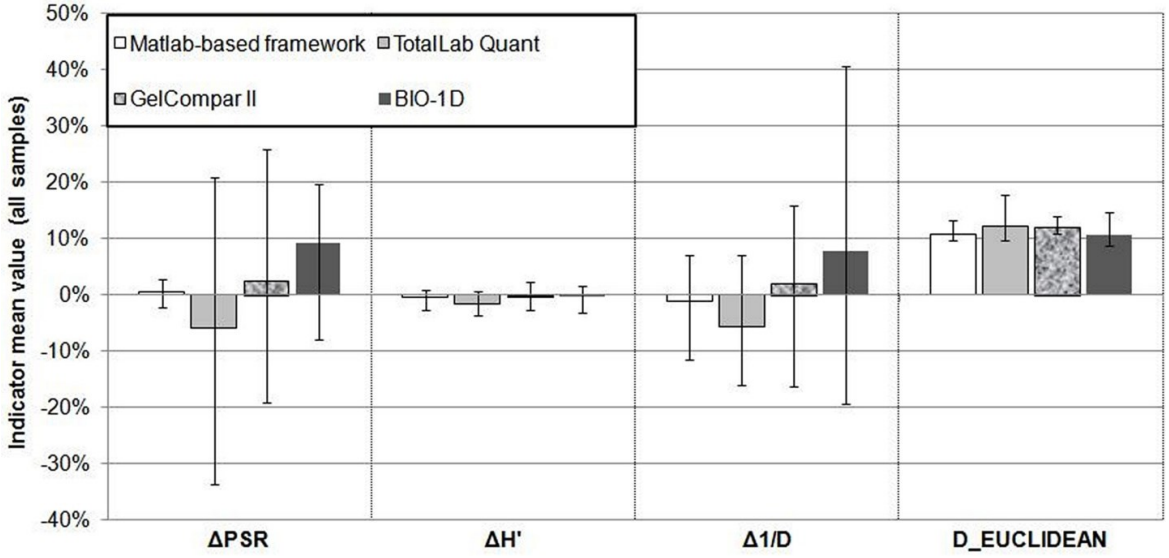

Figure 4-3 : Indicator values obtained using the parameters presented in Tableau 4-2, reported as the mean of all nine samples. Error bars correspond to the maximum and minimum values. All diversity measures were calculated using truncated RADs.

The differences in the results presented in Tableau 4-1 and 4-2 for the three commercial software programs are noteworthy. Taking TotalLab Quant as an example, a ball radius of 20 was never found to be optimal when considering samples individually. Still, this ball size was selected as the best compromise when simultaneously considering all samples. This observation is also true for the similarity level that made it possible to minimize the differences between DGGE and pyrosequencing-based RADs.

The Matlab-based framework was found to yield very stable results over all the samples. Indeed, DGGE-based and true RADs were very close when a similarity level of 98% was chosen—the only exception being $BF_{97\%}$ for which the 97% level gave better results. This framework allowed an accurate extraction of the sample PSRs from the profiles.

Among the indices tested, H' presented a very good match between DGGE-based and true RADs. However, this index was found to be influenced by RAD lengths more than community dominance profiles. Considering that, for the sake of comparison, distribution lengths were forced to be equal for both approaches, it is not surprising that H' exhibited very low variability for all the samples and software programs. As a dominance index, Simpson's 1/D was much more dynamic and is therefore a better indicator than H' to compare software programs. On average, the Matlab-based framework performed better than the other software programs with $\Delta 1/D$ close to ±5% for all samples except $BF_{97\%}$ (-11%). All four software programs yielded a similar average Euclidean distance value around 10-12%.

4.2.4.2 Using PSRs to improve DGGE-based diversity surveys

Based on simulated communities, it was shown that the subunit background percentage, an indicator analogous to PSR, was related to community richness (Loisel et al., 2006). Figure 4-4 shows the relationship between PSRs extracted from DGGE profiles using the Matlab-based framework and true community richness at a 98% similarity level. This similarity level was chosen since it was found to be the one that best corresponded to the extent of DGGE peak clustering. Considering that a relationship between PSR and richness is clearly visible, this parameter can be thought as useful to estimate community diversity from partial DGGE-based RADs.

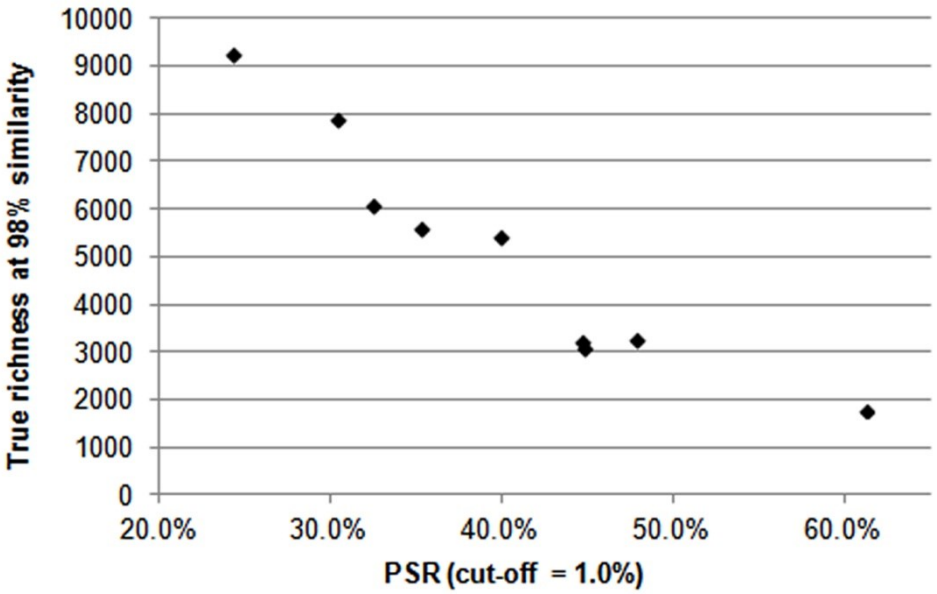

Figure 4-4 : PSRs extracted from the in silico DGGE profiles using the Matlab-based framework compared to true community richness at 98% similarity level.

Calculated solely from the community dominance profile, the diversity indices presented so far are known to be uncorrelated with whole community diversity (Blackwood et al., 2007). Tableau 4-3 illustrates the comparison between richness, Shannon H' and Simpson's 1/D indices calculated from DGGE-based dominance profiles or elongated using the PSRs with the corresponding true indices (untruncated RADs) at a 98% similarity level.

Tableau 4-3 : Deviation of DGGE-based diversity estimates from the indices calculated using the untruncated true RADs. DGGE-based indices were calculated solely from the quantified peaks (with a cut-off of 1.0%) or using the elongated RADs. Results are presented as the average, minimum and maximum of all nine samples, as well as the slope and coefficient of determination generated by linear regression analysis.

Diversity index	Analytical methodology	Relative deviation from true index			Linear regression	
		Average	Min	Max	Slope	R^2
Richness	Peaks quantification	-99.3%	-99.7%	-98.3%	-4.5E-05	0.006
	Elongation framework	4.5%	-10.8%	30.5%	0.93	0.969
Shannon	Peaks quantification	-52.6%	-59.2%	-41.5%	-3.7E-02	0.075
	Elongation framework	2.9%	0.4%	7.1%	1.16	0.992
Simpson's 1/D	Peaks quantification	-88.6%	-95.7%	-74.2%	3.1E-03	0.009
	Elongation framework	3.7%	-3.2%	9.3%	0.94	0.996

4.2.5 **Discussion**

4.2.5.1 Representativeness of DGGE-based dominance profiles

In light of the complete results produced for this paper (TableauA3-1.xlsx), it can be concluded that analytical parameters highly influence DGGE-based diversity surveys of complex bacterial communities. The algorithm used to implement the rolling ball background subtraction method, the chosen ball radius and the way peaks are delimited and quantified all influence the results to a

certain extent. The most influential step is undoubtedly background noise subtraction. Indeed, by sharing the same rolling ball algorithm, TotalLab Quant and GelCompar II behaved similarly while BIO-1D was completely different. Furthermore, indicator values were highly dynamic to ball size modifications for all the software programs. It must be stressed that no ball size fit all samples equally well.

As presented in Tableau 4-1, the closest matching similarity level was not the same for all the samples when DGGE profile background noise was subtracted using rolling ball approaches. This observation mostly implies that it is impossible to know on what basis samples are compared when DGGE profiles are analyzed using automated background subtraction algorithms. The conclusions that can be drawn from DGGE-based diversity studies of soil bacterial communities—at least when profiles are analyzed using the three commercial software programs tested in this publication—are therefore highly limited. It was not possible to identify any relationship between true community diversity and optimal ball radius.

The results produced by the Matlab-based framework that was developed were completely different. Indeed, mean indicator values associated with this methodology were all very close to zero (except for Euclidean distances, discussed later) and generated rather narrow error bars. Furthermore, results were stable over a wide range of cut-off values and consistent throughout all samples. Most importantly, it is the only methodology that made it possible to extract accurate PSRs from DGGE profiles. The ability of this framework to consistently match the samples' true dominance profile at 98% similarity level should not be seen as a coincidence. Using optimized Gaussian PDFs to delimit and quantify peaks was an interesting feature, since it made it possible to determine the bands generated by two almost co-migrating OTUs. However, the consistency of the framework was mostly associated with the manual adjustment of the background profiles that made it possible to treat every sample equally, while rolling ball approaches proved to be highly dependent on how peak superposed in the profiles. It must be acknowledged that this approach was long and challenging at first and thus required some training. Working with *in silico* DGGE gels proved to be a very good way of producing such training sets. Indeed, real DGGE gels shared a lot of similarities with *in silico* profiles, even if they are imperfect and therefore more challenging.

The cut-off value that had to be applied to DGGE-based RADs in order to match true RADs was surprisingly high: 1.0% for the Matlab-based framework. Indeed, more than half of the quantified

peaks are subtracted when using such a high cut-off value. In fact, bands superposition was found to happen quite locally, and background profiles varied irregularly throughout the profile lengths. For the datasets used for this publication, the background was high at the profiles' center but low at their beginnings and ends. Consequently, rather rare OTUs will generate distinct peaks in regions where background is low, while more abundant OTUs migrating in regions of high background will not. These weak bands therefore generated divergences between DGGE-based RADs and true RADs and had to be subtracted.

An important objective of this paper was to evaluate the extent of DGGE bands clustering. From the *in silico* DGGE profiles produced here, when studying complex soil bacterial communities, co-migration events can be seen as the norm rather than the exception. As unambiguously demonstrated by other authors (Schmalenberger and Tebbe, 2003), all DGGE bands can be expected to contain many different OTUs. Working with *in silico* profiles led to the observation that peaks are indeed formed by the addition of one dominant and many rare phylotypes. Even though a high number of OTUs had the exact same calculated T_m and therefore co-migrated on the *in silico* profiles (OTU positions were predicted from the sequences calculated T_m), cases in which two dominant OTUs shared the same T_m were not observed in any sample. It is important to keep in mind that this situation could happen in other samples. For all the samples, it was still possible to identify certain peaks formed by the exact co-migration of many mid-dominant phylotypes, leading to the appearance of a dominant DGGE band. Euclidean distance values presented in Figure 4-3 were mostly driven by the presence of certain important peaks with no corresponding dominant OTUs. For some samples, the presence of a very abundant OTU in the true RADs associated with a DGGE band having a much lower relative abundance also had a significant impact on the resulting Euclidean distances. This happened when two OTUs, both dominant at 100% similarity, were found to cluster when choosing a lower level. DGGE bands corresponding to these OTUs did not cluster on the gel. Considering the lack of resolving power of 16S rRNA gene sequences to identify bacteria at the species level (Rosselló-Mora and Amann, 2001), it is impossible to state unambiguously that these OTUs should be clustered. Still, the Euclidean distances relatively small values clearly indicate that most of the DGGE peaks were associated with a dominant OTU having a quantitatively comparable relative abundance.

It can therefore be concluded that the DGGE peaks and DNA sequences cluster in two different ways. Peaks cluster on the basis of their melting properties more so than on the basis of base-to-

base sequences similarity. Of course, melting properties are linked to sequence composition but a high base-to-base similarity does not guarantee that two sequences will migrate at a similar position, at least in *in silico* DGGE profiles. While different in nature, both processes consistently yielded comparable dominance profiles at 98% similarity level, a value that could change slightly for experimental DGGE gels.

4.2.5.2 Using PSRs to improve DGGE-based diversity surveys

Based on numerical simulations or pyrosequencing studies, a typical soil bacterial community RAD can be confidently described as long (Fierer et al., 2007; Gans et al., 2005; Roesch et al., 2007), rather steep for the most abundant phylotypes and then slowly decreasing toward an asymptotic relative abundance value (doubletons and singletons) (Doroghazi and Buckley, 2008; İnceoğlu et al., 2011; Narang and Dunbar, 2004). The elongation framework developed for this paper aimed to reproduce these characteristics. As presented in Figure 4-4, the ability to accurately extract PSRs from the profiles is a strong prerequisite to estimate true community diversity from DGGE.

The elongation framework presented here was developed using nine samples originating from only three distinct pyrosequencing datasets. The numerous clustering steps involved may have modified the shape of the resulting RADs. By using more datasets, covering many different environments and containing enough reads per sample to reach the plateau of the rarefaction curves, it would be possible to develop a more robust framework adjustable to the many different soil types/environments that researchers may study. Though highly empirical, this rather simple framework proved to be very effective in predicting true community diversity using both the Shannon and Simpson's 1/D indices. As presented in Tableau 4-3, all diversity estimates were accurate at ±10% and highly correlated with true indices at 98% similarity. These values are tremendously better than those produced before RAD elongation, even if these partial distributions were found to be highly representative of true community dominance profiles at 98% similarity.

For some samples, divergence in community richness was found to be higher than ±10%. For some samples, divergence in community richness was found to be higher than ±10%. These differences *per se* were not considered to be a strong shortcoming since the elongation framework did not specifically aim to accurately predict community richness. The usefulness of the framework lies in its capacity to consider true community dominance and bring *H'* and *1/D* indices calculation to

higher and more realistic richness values. Results by Narang and Dunbar (2004), among others, showed that diversity indices are less sensitive to richness at these high values. For 1/D, the elongation framework presented here is somewhat similar to a published methodology (Loisel et al., 2009). These authors proposed the use of a correction factor, also based on the background noise level, to estimate accurate 1/D values from fingerprints. As a dominance index, 1/D was found to be highly sensitive to the OTUs with the highest abundances. Its calculation can therefore be seen as pretty robust to the elongation framework but must rely on an accurate peak quantification step.

As stated by Hill et al. (2003), H' gives more weight than 1/D to rare species and is essentially an intermediate between community richness and the Simpson index. This index is therefore less affected by the peak quantification step than 1/D but requires an acceptable estimation of community richness in order to be accurate. In the course of developing the elongation framework, it was observed that the accuracy of H' depended on the number of rare species more than on the trajectory of the RADs. Therefore, as long as the predicted richness was fairly accurate, using one or another abundance distribution for the RAD elongation step did not change H' by more than ±10%, even though the correspondence between the elongated and true RADs was not very good. It will therefore be important to validate that the proposed elongation framework is able to yield acceptable community richness predictions on real—and more challenging—samples before using these values. However, it will always be possible to adapt the model parameterization to different situations (primer pairs, studied environments, etc.) whenever necessary.

A question that remains is the usefulness of traditional DGGE-based diversity surveys that rely solely upon the visible peaks. Tableau 4-3 clearly shows that these surveys strongly underestimate the diversity of soil bacterial communities. More importantly, from the diversity indices presented as supplementary material (Tableau A3-3), it has been observed that these studies are susceptible to erroneous ecological conclusions, often showing no differences between samples when important ones exist or sometimes predicting the opposite.

In conclusion, the framework presented in this paper proved to be very successful at estimating true community diversity for all nine *in silico* DGGE profiles analyzed. Though only Shannon and Simpson's 1/D indices were evaluated, the very good correspondence between all the DGGE-based and true community RADs at 98% similarity leads to the hypothesis that the framework will

accurately estimate any diversity index influenced by community structure more than richness. Imperfect in nature, experimental DGGE gels are much more challenging to analyze than the image synthesized here. Consequently, when working with experimental results, deviations from true community diversity can be expected to be higher. At the moment, it is not possible to provide a quantitative estimate of the expected deviation. Potentially important biases in sequencing datasets, often linked to sequence GC content, were reported (Dohm, Lottaz, Borodina, and Himmelbauer, 2008; Jaenicke et al., 2011; Pinard et al., 2006; Pinto and Raskin, 2012). Until these issues are resolved, whether or not next-generation sequencing platforms offer a more solid ground than DGGE to quantitatively estimate the diversity of soil bacterial community remains an open question that deserves further attention.

4.2.6 **Acknowledgements**

The authors acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada. The CIRAIG would also like to thank its industrial partners for their financial support: ArcelorMittal, Bell Canada, Bombardier, Cascades, Mouvement des caisses Desjardins, Groupe Electricite de France/Gaz de France, Eco Entreprises Quebec, Hydro-Quebec, Johnson & Johnson, Groupe Louis Vuitton Moët Hennessy, Michelin, Nestlé, Recyc-Quebec, Rio Tinto Alcan, RONA, Societe des Alcools du Quebec, Solvay, Total, Umicore, and Veolia Environment.

CHAPITRE 5 VALIDATION DE LA MÉTHODOLOGIE D'ÉLONGATION DES RAD DGGE SUR DES ÉCHANTILLONS DE SOL CONTAMINÉS AU PENTACHLOROPHÉNOL

Cette section est centrée autour du second manuscrit intitulé « *Can DGGE and Ion Torrent sequencing yield similar quantitative conclusions when comparing the diversity of soil bacterial communities?* » (section 5.2) soumis à la revue *FEMS Microbiology Ecology* (soumis le 23 avril 2013).

5.1 Lien entre le premier et le second manuscrit

Le premier manuscrit a permis la validation théorique de la méthodologie d'élongation des RAD issues des gels DGGE. Par contre, la validité de cette approche doit encore être montrée expérimentalement. En effet, les gels DGGE expérimentaux sont imparfaits et représentent donc un défi plus important que les profils *in silico*. Avant de l'utiliser, il est donc impératif de tester cette approche sur des échantillons réels. Le second manuscrit vise donc à comparer les estimés de diversité produits à partir du DGGE et à partir de l'Ion Torrent, une plateforme NGS. Pour ce faire, deux sols de texture différente sont divisés en quatre sous-échantillons. Pour chaque sol, trois de ces échantillons sont contaminés à différentes concentrations en PCP (100, 900 et 3000 mg/kg).

En analysant les mêmes échantillons à partir de deux approches différentes, ce manuscrit vise également à identifier les sources de distorsion inhérentes aux deux techniques et à proposer des moyens permettant de réduire leur importance. Puisque la diversité des communautés bactériennes des sols étudiés n'est pas connue, il ne sera pas possible de conclure quant à la supériorité de l'une ou l'autre des approches utilisées. Par contre, l'existence d'une relation linéaire significative entre la diversité estimée à partir du DGGE et de l'Ion torrent permettra de conclure que l'une et l'autre de ces approches ont la capacité de produire des résultats similaires lorsqu'utilisées à des fins comparatives.

5.2 "Can DGGE and Ion Torrent sequencing yield similar quantitative conclusions when comparing the diversity of soil bacterial communities?"

5.2.1 Abstract

Denaturing gradient gel electrophoresis (DGGE) is a technique that has been used extensively to study soil bacterial communities. Considering only the most abundant phylotypes, this technique is known to produce diversity estimates uncorrelated to true community diversity. Recently, a framework to accurately quantify the diversity of these communities using DGGE was published. This paper aims to verify if this framework and next-generation sequencing platforms can yield similar diversity estimates when working on real samples. To do so, the impact of pentachlorophenol on bacterial diversity was quantified using both DGGE profiling and Ion Torrent sequencing. Using two different techniques to study the same communities made it possible to find biases for both methodologies. For DGGE, localized gel saturation was identified as the most important bias. For the Ion Torrent datasets, many previously described biases (sampling depth, sequencing errors and GC content considerations) proved to influence the diversity estimates in a sample-specific way. Reducing the extent of the different sources of bias allowed finding significant correlations between the diversity estimates produced by DGGE and Ion Torrent. The results are promising since two very different techniques were able to yield similar conclusions when comparing the diversity of soil bacterial communities.

5.2.2 Introduction

As global diversity decreases and the pressure on natural ecosystems increases (Butchart et al., 2010), environmental microbiologists now require tools to routinely assess, compare and follow soil bacterial community diversity trends. These tools must be fast, reproducible, accessible and, most of all, produce environmentally relevant diversity estimates. Even if these highly diverse and complex communities (Roesch et al., 2007) can be studied in great depth using next-generation sequencing techniques (Simon and Daniel, 2011), the methods remain costly and inaccessible to many researchers.

Denaturing gradient gel electrophoresis (DGGE) is a widespread technique that was used to compare and differentiate soil bacterial communities in numerous situations. Among others, it was successfully used to monitor temporal variations in bacterial communities (Lopes et al., 2011),

differentiate between agricultural practices (Lopes et al., 2011; Rahman, Okubo, Sugiyama, and Mayland, 2008) and evaluate the impact of inorganic (Lorenz et al., 2006; Xie et al., 2011) and organic (Martí et al., 2011; Zielezny et al., 2006) contaminants on soil bacterial community diversity. However, at their sensitivity threshold, it was demonstrated that fingerprinting techniques such as DGGE yield estimates that are not correlated with the true diversity of the communities (Blackwood et al., 2007). Recently, a new framework that makes it possible to generate accurate diversity estimates of soil bacterial communities from DGGE was published (Lalande, Villemur, and Deschenes, 2013). This framework associates the visible DGGE bands to dominant species and links the background noise found under the profiles to the many subdominant phylotypes. Indeed, using the peak-to-signal ratio (PSR) parameter, partial rank-abundance distributions (RADs) drawn from DGGE peaks quantification are corrected and then elongated in order to reproduce the true community RADs. Developed from *in silico* DGGE profiles synthesized from 454 pyrosequencing datasets of soil bacterial communities (Nacke et al., 2011), this methodology made it possible to estimate both the Shannon H' and Simpson's 1/D index (Magurran, 2004) with a ±10% precision for the nine samples that were studied. Community richness was less precise but still ranged between -10 to 30% of the expected values. The framework has yet to be validated on real samples.

Even if next-generation sequencing technologies are considered to be the most powerful tools currently available to study soil bacterial communities, many potential limitations were recently associated with the techniques. First, an insufficient sequencing effort could prevent the recovery of the full extent of the diversity (Nacke et al., 2011). Also, potential distortions associated with sequence GC content were identified. Indeed, it was demonstrated that 454 pyrosequencing platforms have more difficulty producing long high-quality reads for the amplicons with high GC content (Jaenicke et al., 2011). In their dataset, Pinto and Raskin (2012) also found that the number of reads per sample was related to the samples' average GC content. Based on this finding, they encouraged more research to determine whether or not GC bias can affect the relative abundances of OTUs and thus cause intra-sample bias. Finally, sequencing errors were reported to artificially inflate all richness-based metrics (Will et al., 2010).

The primary objective of this paper was to validate the DGGE RADs elongation framework on pentachlorophenol (PCP) contaminated soils. The second objective was to determine whether the potential biases in sequencing datasets could cause important intra-sample distortions in the

relative abundances of OTUs.

5.2.3 Methods

5.2.3.1 Preparation of spiked soils

About 2 kg of sandy loam (LM) and sandy soil (S) were sampled near Montréal (QC, Canada), immediately sieved (2-mm mesh) and stored at 4°C until use. The particle size distribution was determined according to American Society for Testing and Materials (ASTM) methods D1140-00 (ASTM International, 2006) and D422-63 (ASTM International, 2007a). The soils were classified according to the USDE classification system (sand [2mm-50µm], silt [50-2µm], and clay [<2µm]). Total carbon (TC) was measured by combustion with an induction furnace (LECO Corporation, St-Joseph, MI) according to the ASTM method D2974-07a (ASTM International, 2007b). Total inorganic carbon (TIC) was performed with a phosphoric acid treatment followed by an infrared determination. Total organic carbon (TOC) was calculated from the difference between TC and TIC. Soil pH was determined with distilled water using potentiometry (Orion Surfow semi-micro Ross combination pH electrode) according to ASTM method D4972-01 (ASTM International, 2007c) with a soil to water ratio of 1:2. Soil granulometry, total organic carbon content and pH are presented in Tableau 5-1. Each soil was divided into four subsamples; one was kept clean and the three others were contaminated with pentachlorophenol (PCP) at concentrations of 100, 900 and 3,000 mg (kg dry wt soil)$^{-1}$.

Tableau 5-1 : Physico-chemical properties of the soils used in the study.

Soil	% sand	% loam	% clay	Total organic carbon (%)	pH
LM	58	24	18	2.8 ± 0.4	7.39 ± 0.03
S	98	2	0	2.2 ± 0.1	6.68 ± 0.03

Contamination was performed by dissolving the appropriate amount of PCP in 5 mL of dichloromethane in empty 250 mL amber bottles. The bottles were gently shaken during dichloromethane volatilization in order to recover PCP on the bottle walls. After complete dichloromethane volatilization, soil samples (40 g) were added to the bottles and mixed at room temperature in a tumble action agitator for 24 hours in order to recover the PCP from the bottle walls. Samples were humidified with distilled water to 60% of the soils' water holding capacity

and finally manually mixed. Soil samples were kept in the dark at room temperature for 28 days prior to DNA extraction.

5.2.3.2 DNA extraction and PCR amplification

DNA extractions were carried out using a slightly modified version of a previously described bead-beating method (Bürgmann, Pesaro, Widmer, and Zeyer, 2001). Briefly, soil (300 mg) and 0.1 mm diameter zirconia-silica beads were mixed in a 1:1 ratio (w/w). Bead-beating was done at maximum speed (2800 rpm) for 90 seconds with a Mini BeadBeater-8 (BioSpec Products Inc, Bartlesville, OK, USA) in 650 µl of extraction buffer composed of 0.2% (w/v) hexadecyltrimethylammonium bromide (CTAB), 1 mM dithioteitrol (DTT), 0.2 M sodium phosphate buffer (pH 8), 0.1 M sodium chloride and 50 mM ethylenediaminetetraacetic acid (EDTA). Purification steps involved polyvinylpolypyrrolidone (PVPP) spin columns, incubation/centrifugation in a 20% (w/v) polyethylene glycol and 2.5 M NaCl solution and phenol:chloroform:isoamyl alcohol extraction (Thakuria, Schmidt, Mac Siurtain, Egan, and Doohan, 2008). DNA was recovered by ethanol precipitation (Sambrook, Fritsch, and Maniatis, 1989). Purified DNA was quality-checked using a 1% (w/v) agarose gel and quantified with a TECAN GENios fluorometer (Tecan Group Ltd., Männedorf, Switzerland). Samples were stored at -20°C.

For both DGGE and Ion Torrent sequencing, PCR amplification targeted the V3 region of the bacterial 16S rRNA gene with primers 341f (5'-CCTACGGGAGGCAGCAG-3') and 534r (5'-ATTACCGCGGCTGCTGG-3') (Muyzer, de Waal, and Uitterlinden, 1993). From the results produced by the ProbeMatch tool (Cole et al., 2005) with the primers, amplicons of bacterial origin of lengths between 133 and 178 base pairs (bp) were to be expected with major peaks at 136 and 161 bp and minor ones at 138, 141 and 156 bp. The DGGE forward primer also contained a 40 nt GC clamp at its 5' end (5'-CGCCCGCCGCGCGCGGCGGGCGGGGCGGGG GCACGGGGGG-3') (Muyzer et al., 1993). Sequencing primers contained the Ion Torrent adaptor A followed by a 10 nt sample specific multiplex identifier (MID) (forward) or adaptor trP1 (reverse) at their 5' end. Ion Torrent adaptors and MID sequences are listed in Tableau A4-1. Amplifications were conducted in 50-µL volumes using the 2X Phusion® High-Fidelity PCR Master Mix (Thermo Fisher Scientific Inc., Waltham, MA, USA). PCR mixtures contained 200 µM of each dNTP, 3.5 mM of $MgCl_2$, 0.6 g L^{-1} of bovine serum albumin (BSA), 0.2 µM of each primer and 5 ng of the extracted DNA. Amplification was carried out with a Techne Genius Thermocycler (Bibby Scientific Ltd.,

Staffordshire, UK). After an initial two-minute denaturation step at 94°C, a total of 40 amplification cycles were carried out, consisting of 15 seconds denaturation at 94°C, 15 seconds annealing and 15 seconds elongation at 72°C. The amplification process was concluded with a final 10-minute elongation step at 72°C. To increase the specificity and yield of the amplification process (Korbie and Mattick, 2008), the annealing temperature was decreased from 67.5 to 62.5°C during the first 10 cycles and then kept at 62.5°C for the remaining 30 cycles. The PCR conditions were optimized for the sequencing primers that had a strong tendency to dimerize. Dimerization was not a problem for DGGE primers but the same PCR conditions were used. Amplification specificity and yield was checked on a 1.8% (w/v) agarose gel using an EZ Load™ Precision Molecular Mass Ruler (Bio-Rad Laboratories Inc., Hercules, CA, USA). For DGGE, all amplifications were carried out in triplicate in order to assess the variability attributable to PCR amplification and DGGE profiling.

5.2.3.3 DGGE profiling and analysis

DGGE profiling was conducted using a D-Code system (Bio-Rad Laboratories Inc., Hercules, CA, USA). The DNA extracted from each sample was PCR-amplified three times and three different DGGE gels were produced. The DNA of all the samples was amplified three times The 8% (w/v) acrylamide gel was cast into 16 x 16 cm plates with 1 mm thick spacers and composed of a 32.5-72.5% denaturing gradient, 100% corresponding to 7 M urea and 10% (v/v) deionized formamide. Approximately 400 ng of amplified DNA was loaded into the gel for each sample. A blank PCR sample was also added at the beginning of each gel. Electrophoresis was run at 60°C and 60V for 18 hours. The gel was stained for 30 minutes in a fresh solution of SYBR® Gold nucleic acid gel stain (Life Technologies, Carlsbad, CA, USA). After staining, the gel was photographed using a Quantum ST4 gel documentation system (Vilber Lourmat, Marne-la-Vallee, France).

DGGE profiles analysis was done using a newly developed Matlab-based program (The MathWorks Inc., Natick, MA, USA). The background noise of the profiles was manually subtracted, since the methodology yields results that are much more consistent than the widespread rolling ball approaches (Lalande, Villemur, et al., 2013). Banding patterns were then analyzed using Gaussian probability density functions to represent the peaks. Partial rank-abundance distributions (RADs) originating from peaks quantification were first truncated using a cut-off value (1.0%) that enabled the removal of the peaks that were too small to be considered representative of the communities' true dominance profiles. Using the peak-to-signal ratio (PSR)—

a parameter extracted from the DGGE profiles and related to community richness (Loisel et al., 2006)—truncated RADs were elongated in order to include in the diversity assessment the species that were not abundant enough to produce a visible band on the gel. Designed to infer from the DGGE profiles the RADs that would be produced by sequencing datasets containing 35,000 reads, this elongation framework yielded diversity indices that were highly similar to the clustering of the sequencing datasets at the 98% similarity level. Please refer to the original publication for a more complete description of the entire profile analysis and elongation (Lalande, Villemur, et al., 2013).

5.2.3.4 16S rRNA gene amplicon sequencing and bioinformatic analysis

Prior to emulsion PCR, amplification products intended for sequencing were purified on agarose gels using a GenElute gel extraction kit (Sigma-Aldrich, St-Louis, MO, USA) and quantified using a PicoGreen double-stranded DNA quantitation assay (Life Technologies, Carlsbad, CA, USA). The amplification products from the different samples were pooled in an equimolar ratio. Bacterial 16S rRNA gene amplicons were sequenced at the National Research Council of Canada in Montréal, Québec. As described by Yergeau et al. (2012), the emulsion PCR and sequencing was done using an Ion OneTouch 200 template kit (Life Technologies, Carlsbad, CA, USA), OneTouch and OneTouch ES instruments (Life Technologies, Carlsbad, CA, USA) and an Ion Torrent personal genome machine (PGM). Even though it is newer and therefore less frequently used than other sequencing platforms, Ion Torrent sequencing produced community structure and composition that were almost interchangeable with those produced by 454 pyrosequencing (Yergeau et al., 2012). The sequences were deposited in the NCBI sequence read archive (NCBI, 2012) under the accession number SRR833619.

Sequencing dataset processing was done in such a way that only the sequences containing both primers were retained. Initial processing was carried out using Matlab. Since the general quality of the sequences dropped at a length of ~100 bases, very few sequences produced hits with the reverse primer. To overcome this limitation, the reverse primer was trimmed to allow 7 mismatches (over 17 nt). For the barcodes and the forward primer, only one mismatch was tolerated. The remainder of the dataset processing was done using Mothur (Schloss et al., 2009). First, trimmed amplicons with unresolved bases (N's), homopolymers longer than 7 or an average quality score under 20 were discarded. Datasets were aligned, and amplicons outside the general alignment space or the expected length range were disregarded. Aligned datasets were then screened for chimeras using

UCHIME (Edgar, Haas, Clemente, Quince, and Knight, 2011). For all the samples, amplicons identified as chimeras represented less than 2.5% of all the sequences. Since no denoising program seems to have been adapted to the Ion Torrent PGM, sequencing errors were reduced using the Mothur implementation of a pseudo-single linkage clustering algorithm (Huse et al., 2010). As recommended, sequences with less than 2% dissimilarity (3 bases) were clustered to more abundant OTUs. Finally, clustering was done using an average neighbour algorithm. This methodology (pseudo-single linkage followed by average linkage clustering) proved to accurately predict expected OTUs for communities of known taxonomic composition (Huse et al., 2010). Contrary to what is generally done, final clustering was carried out at 98% similarity as, on *in silico* profiles, this level consistently matched the DGGE banding patterns (Lalande, Villemur, et al., 2013).

5.2.3.5 Diversity quantification

Diversity estimations produced by the elongation framework and sequencing dataset clustering at the 98% similarity level were compared through three diversity indices. Community structure was compared using the Shannon and Simpson indices. Considering the arguments of Jost (2006), diversities were reported in the form of an effective number of species. Therefore, the Shannon index was calculated as $\exp^{H'}$ and Simpson was calculated as 1/D. The Shannon index is reported to give more weight to rare OTUs as compared to the common ones, while the Simpson index is more affected by dominants (Hill et al., 2003). The information conveyed by both indices was therefore seen as complementary. Community richness was also compared. For the Ion Torrent results, richness corresponded to the actual number of observed OTUs after the clustering process. For DGGE, this value corresponded to the length of the elongated RADs. All calculations were performed using the PAST software (Hammer et al., 2001).

5.2.3.6 Identifying potential biases in DGGE and sequencing datasets

A first analysis of the results made it possible to identify potential biases in both DGGE and sequencing-based diversity estimates. For DGGE, it was hypothesized that peak intensities may not vary linearly based on the amount of DNA that generated the bands. Different amounts (25 to 400 ng) of *Escherichia coli* 16S rRNA gene PCR amplifications were loaded on a gel. Also, the DNA amplification and DGGE profiling steps were done again (in triplicate), this time by loading approximately 200 ng of PCR product per well. It must be acknowledged that DNA extraction had to be performed again for S_{3000} since the initial extract had been completely used. Between both

extractions, the sample was kept in the dark at -20°C. The methodology for PCR amplification, DGGE profiling and profile analysis was also further optimized. Indeed, amplifications were carried out using HPLC purified primers instead of standard purity ones. Also, the $MgCl_2$ concentration was reduced from 3.5 to 2.5 mM, the length of the denaturation, annealing and elongation steps was reduced from 15 to 10 seconds and the number of PCR cycles was reduced from 40 to 35.

Biases in sequencing datasets were associated with sampling depths, sequencing errors and GC content considerations. The impact of the sampling depth was investigated using the elongated DGGE-based RADs (200 ng $well^{-1}$). Since the DGGE RADs elongation framework produced distributions containing 35,000 simulated "reads", they were randomly sampled without replacement using Matlab. Ten different sampling runs were conducted, each time taking between 100 and 30,000 reads. The diversity indices computed from these partial distributions were compared to the indices calculated from the complete elongated RADs. Based on these results, it was possible to derive sample-specific correction factors to approximate the diversity estimates that would have been calculated at a sampling depth of 35,000 reads. The diversity indices are referred to as *Ion Torrent (corrected)*. All of the sequencing datasets were also normalized to the same size (5,000 reads) for all samples in order to yield more comparable indices. This normalization was done by sampling the sequencing datasets as described above.

The potential quantitative impact of sequencing errors was studied by reapplying the previously described pseudo-single linkage clustering algorithm (Huse et al., 2010) with a lower or higher dissimilarity level (2 or 4 bases instead of 3). Finally, bias linked to GC content was qualitatively investigated by reprocessing the initial Ion Torrent dataset as described above (trim primers and remove sequences with homopolymers longer than 6 bases) but without using any quality or length filter. All the sequences were analyzed in order to determine their length, GC percentage and average phred quality score.

5.2.4 Results

5.2.4.1 Assessing the impact of PCP on soil bacterial community diversity with DGGE and Ion Torrent sequencing

A visual inspection of the clean and contaminated DGGE profile samples shows that PCP changed the structure of the soil bacterial communities (Figure 5-1). These structural modifications resulted in diversity decreases for all the contaminated samples when compared to their clean counterparts. Even though the quantitative index values were different for DGGE and Ion Torrent, both techniques showed similar trends. Interestingly, the diversity of LM_{3000} and S_{3000}, the most contaminated samples, is higher than less contaminated ones. This situation could arise if PCP exerts a strong toxic effect on bacterial communities, hampering the species that were able to tolerate lower PCP concentrations to proliferate. Since these samples are quite similar to their clean counterparts, especially for S_{3000}, it is also possible that the profiles were mostly constructed with DNA belonging to dead samples (Marschner et al., 2001).

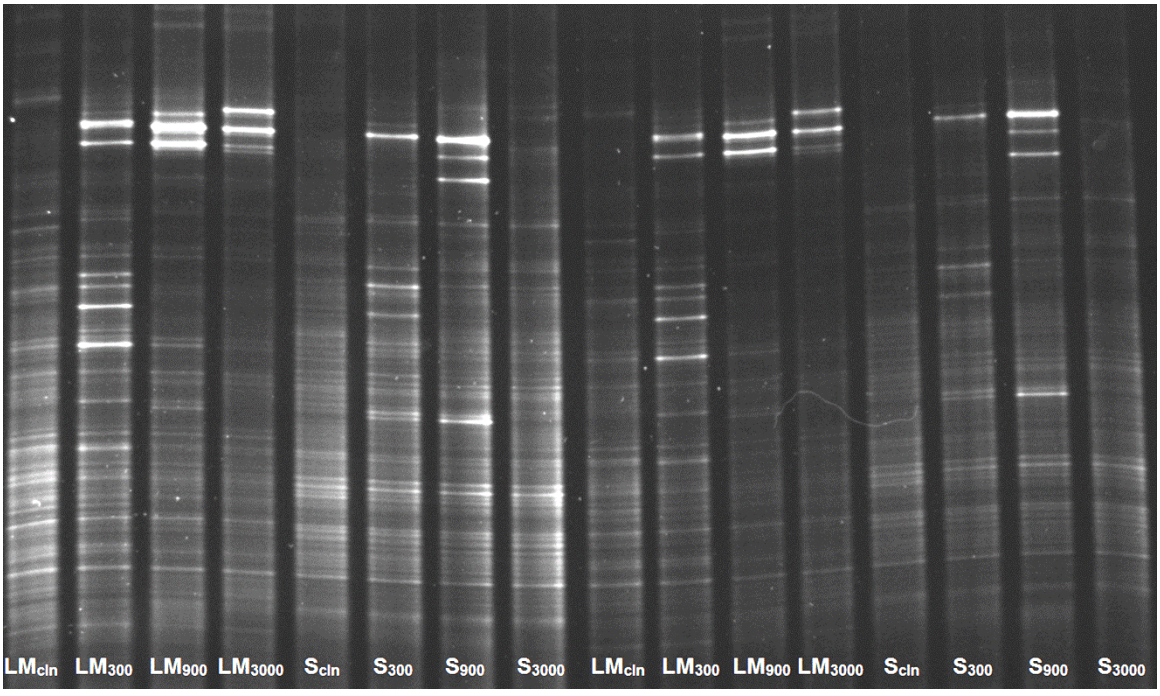

Figure 5-1 : DGGE profiles of the 8 samples analyzed. The DNA loaded on the gel was extracted and amplified from a sandy loam (LM) or sandy soil (S). Samples were either clean (cln) or contaminated with pentachlorophenol (PCP) at concentrations ranging from 300 to 3,000 mg (kg dry wt soil)$^{-1}$. Left: 400 ng of PCR product per well. Right: 200 ng of PCR product per well.

Although DGGE and Ion Torrent sequencing yielded similar index values in some cases, most diversity estimates were found to diverge to varying extents (Figure 5-2). Most of the DGGE 1/D estimates are higher than their sequencing counterparts, indicating differences in the dominance profiles drawn from both techniques. The picture is more complex for $\exp^{H'}$ and OTU richness, since both soils show different patterns. Indeed, most of the DGGE estimates are higher than the sequencing results for soil LM, but the opposite is observed for soil S. Of course, seeing as two very different techniques were compared, some divergences were to be expected. However, considering the good agreements between the diversity indices calculated using the DGGE RADs elongation framework and DNA sequences clustering at 98% when working on *in silico* DGGE profiles, such significant divergences were somewhat surprising. The patterns of the observed discrepancies led to the hypothesis that, apart from the fact that two different techniques were used, both datasets were biased to a certain extent.

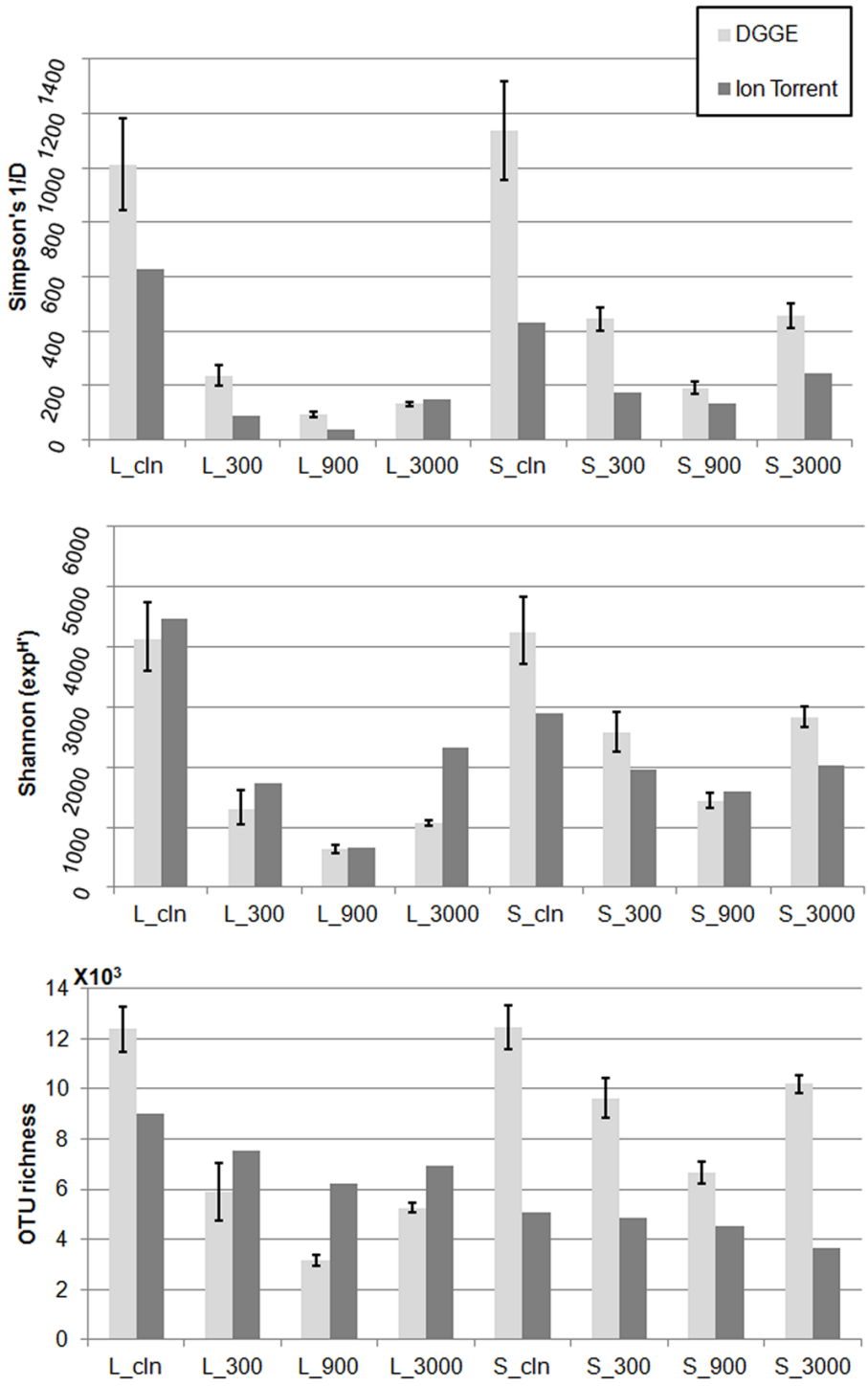

Figure 5-2 : Diversity indices estimated using DGGE or Ion Torrent sequencing. The error bars correspond to the 95% confidence intervals that were calculated from the DGGE gel triplicates. DGGE diversity estimates were calculated from the RADs elongation framework presented in Lalande et al. (2013). Ion Torrent sequencing datasets were clustered at the 98% similarity level using Mothur (Schloss et al., 2009).

5.2.4.2 Biases in DGGE results

Almost all DGGE 1/D values were higher than their sequencing counterparts. Among the explanations related to the sequencing datasets, these types of results were to be expected if the brightest peaks on the DGGE profiles are hampered from reaching their full intensity. It was hypothesized that DGGE gels could be locally saturated by DNA or dye and that the peak intensities did not vary linearly according to the amount of DNA. Figure 5-3 presents the relative peak intensity when loading different amounts of PCR product originating from *E.coli*. An almost linear relationship was found when loading into a well less than 75 ng ($R^2 = 0.89$). Beyond 75 ng (for a single band), the peak intensity begins to saturate.

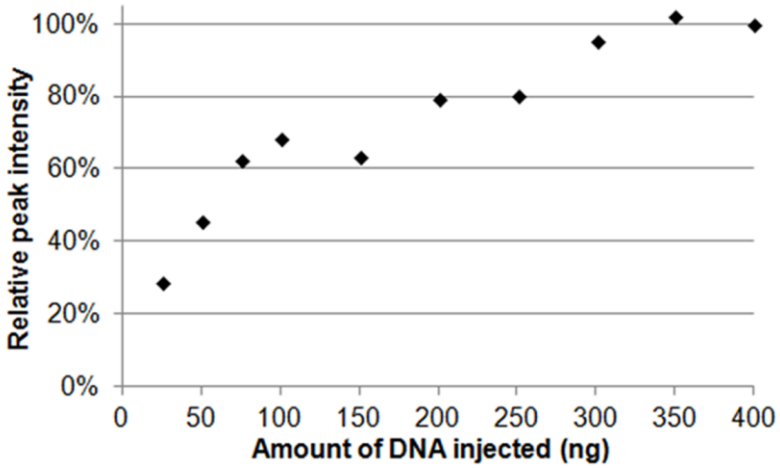

Figure 5-3 : Intensity of the DGGE band produced by loading different amounts of a PCR product amplified from the genomic DNA of *Escherichia coli*. All PCR products came from the same aliquot. Resulting intensities, calculated as the area under the peaks, are reported as a fraction of the intensity corresponding to a loading of 400 ng.

For the DGGE-based (200 ng well^{-1}) diversity indices presented as supplementary material (Tableau A4-2), loading lower amounts of PCR product yielded lower 1/D values for all samples except S_{3000}. This exception can be explained by the fact that different DNA extracts were used to produce the 400 and 200 ng well^{-1} profiles; the brightest band of the initial S_{3000} profile is not visible in the 200 ng well^{-1} gel (see Figure 5-1). For the subsequent analysis, the diversity indices corresponding to S_{3000} (200 ng well^{-1}) were therefore estimated from the 400 ng well^{-1} indices. Generally speaking, loading lower amounts of PCR product led to 1/D values closer to the Ion Torrent estimates. More influenced by rare species than abundant ones, $\exp^{H'}$ and OTU richness

estimates were found to be less affected by the amount of DNA loaded into the wells than 1/D.

5.2.4.3 Biases in Ion Torrent sequencing results

Tableau 5-2 presents the number of reads that match both primers contained within the sequencing datasets before and after processing. The achieved depth was different for all the samples and particularly low for soil S. The proportion of sequences matching with both primers but having a mean quality score under 20 was also different for all the samples. This proportion was higher for the clean and 300 mg PCP kg^{-1} samples when compared with the most contaminated ones.

Tableau 5-2 : For all the samples, number of reads matching both primers before and after processing. The deletion percentage corresponding to the eight samples are also presented.

Sample	After processing	Before processing	Deletion %
LM_{cln}	13,333	37,835	65
LM_{300}	14,069	35,852	61
LM_{900}	15,081	27,589	45
LM_{3000}	11,144	26,894	59
S_{cln}	7,254	22,400	68
S_{300}	7,494	22,329	66
S_{900}	7,277	17,566	59
S_{3000}	5,126	13,327	62

Figure 5-4 presents the evolution of 1/D, $exp^{H'}$ and OTU richness when sampling depth is increased. Considering these curves, it is obvious that the Ion Torrent diversity indices underestimate the diversity of the studied soil bacterial communities. Indeed, even at 35,000 reads, $exp^{H'}$ and especially OTU richness did not reach their plateau values. On the contrary, 1/D values stabilized between 10,000 and 15,000 reads—a value that was not reached for all samples. It is important to note that the bias linked with sampling depth is sample-dependent. The samples showing high diversity and low community dominance required more reads than less diversified ones to reach a given index accuracy. These results show the importance of normalizing the sequencing datasets to the same size prior to comparing diversity indices, especially for OTU richness and $exp^{H'}$. The diversity indices that were produced by two normalizing frameworks (normalized at 5,000 or at 35,000 reads using the DGGE RADs) are presented as supplementary material (Tableau A4-2).

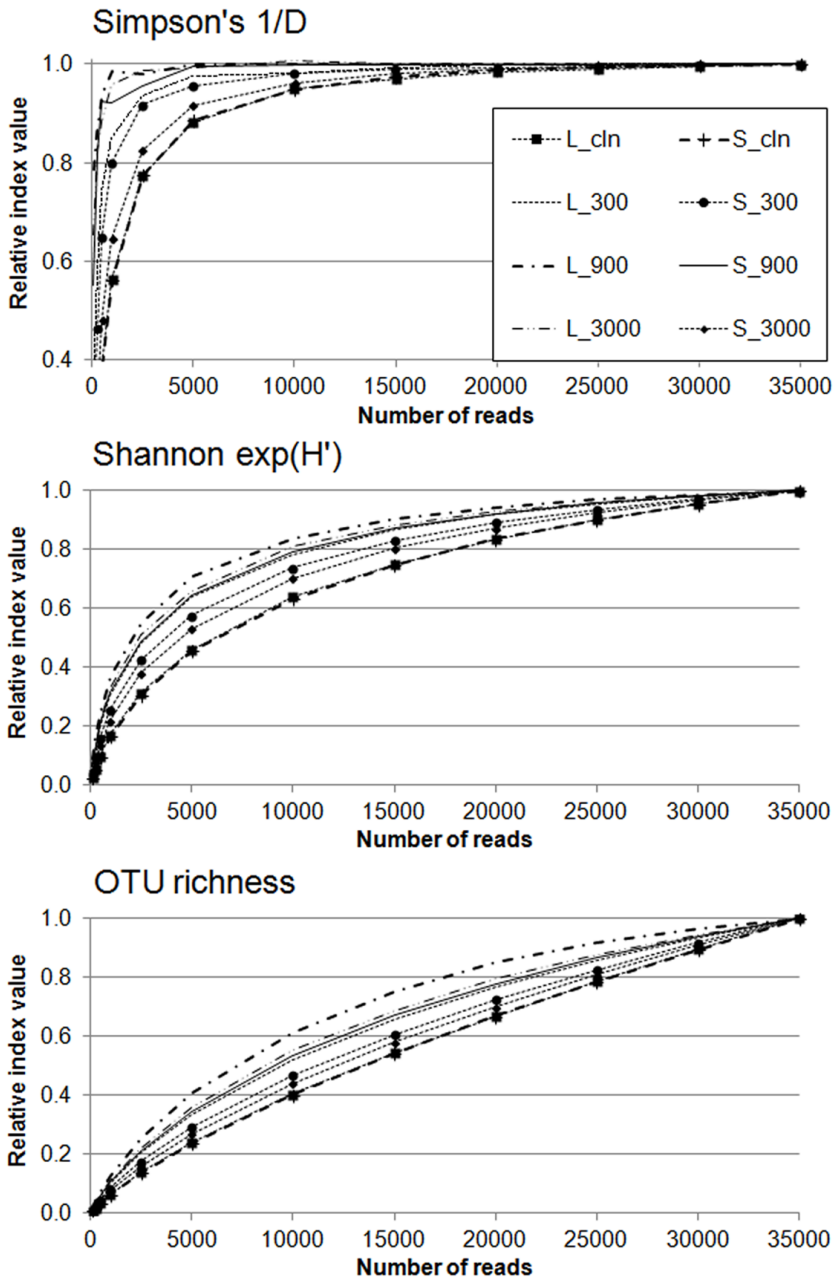

Figure 5-4 : The elongated DGGE RADs, each containing 35,000 reads, were randomly sampled without replacement at various depths ranging from 100 to 30,000 reads. The impact of sampling was quantitatively assessed by dividing the diversity indices calculated at a certain depth by the values calculated at a depth of 35,000. Since DGGE gels were run in triplicate, results were averaged for every sample.

Besides sampling depth, sequencing errors also have the potential to influence diversity estimates. Tableau 5-3 presents the results obtained when applying the pseudo-single linkage clustering

algorithm (Huse et al., 2010) with a dissimilarity of 2 or 4 bases instead of 3. Interestingly, this process influenced community dominance (through 1/D) a lot more than richness. Especially for 1/D, the impact of increasing or reducing the dissimilarity level is not linear and sample-dependent. Changing the dissimilarity level impacted the 1/D values of soil LM samples more than soil S samples. These differences could be explained by the sampling depth achieved for soil S samples (Tableau 5-2), since fewer sequences also mean fewer errors (in absolute values). These observations could indicate that by processes different from undersampling, sampling depth may have more influence on the 1/D values than Figure 5-4 suggests. Although not shown in Tableau 5-3, the pseudo-single linkage clustering identified proportionally more "potentially erroneous" sequences in the contaminated sample datasets. Considering that these datasets were dominated by longer sequences (~160 bases instead of ~136 bases), it can be hypothesized that many sequencing errors were located at the end of the sequences, where the per-base quality scores were relatively low.

Tableau 5-3 : Potential influence of sequencing errors on the indices calculated from sequencing datasets. The values correspond to the relative index variation (in %) when the pseudo-single linkage clustering algorithm (Huse et al., 2010) is applied using a dissimilarity of 2 or 4 bases instead of 3.

Sample	Dissimilarity of 2 bases			Dissimilarity of 4 bases		
	1/D	$exp^{H'}$	OTU richness	1/D	$exp^{H'}$	OTU richness
LM_{cln}	69	24	6	-34	-22	-8
LM_{300}	35	32	7	-25	-28	-10
LM_{900}	48	36	3	-35	-38	-12
LM_{3000}	51	28	6	-30	-27	-9
S_{cln}	42	22	7	-24	-19	-8
S_{300}	25	22	6	-21	-21	-8
S_{900}	39	26	5	-27	-25	-8
S_{3000}	45	22	7	-20	-17	-7

The relationship between sequence length or GC content and quality was further investigated. The upper part of Figure 5-5 presents the relationship between the sequences' average quality and trimmed length. Then, extracting the sequences corresponding to bins of particularly low deletion percentages, amplicons were ordered by increasing GC content values in order to determine how GC content could influence average quality. This process was carried out individually for each length.

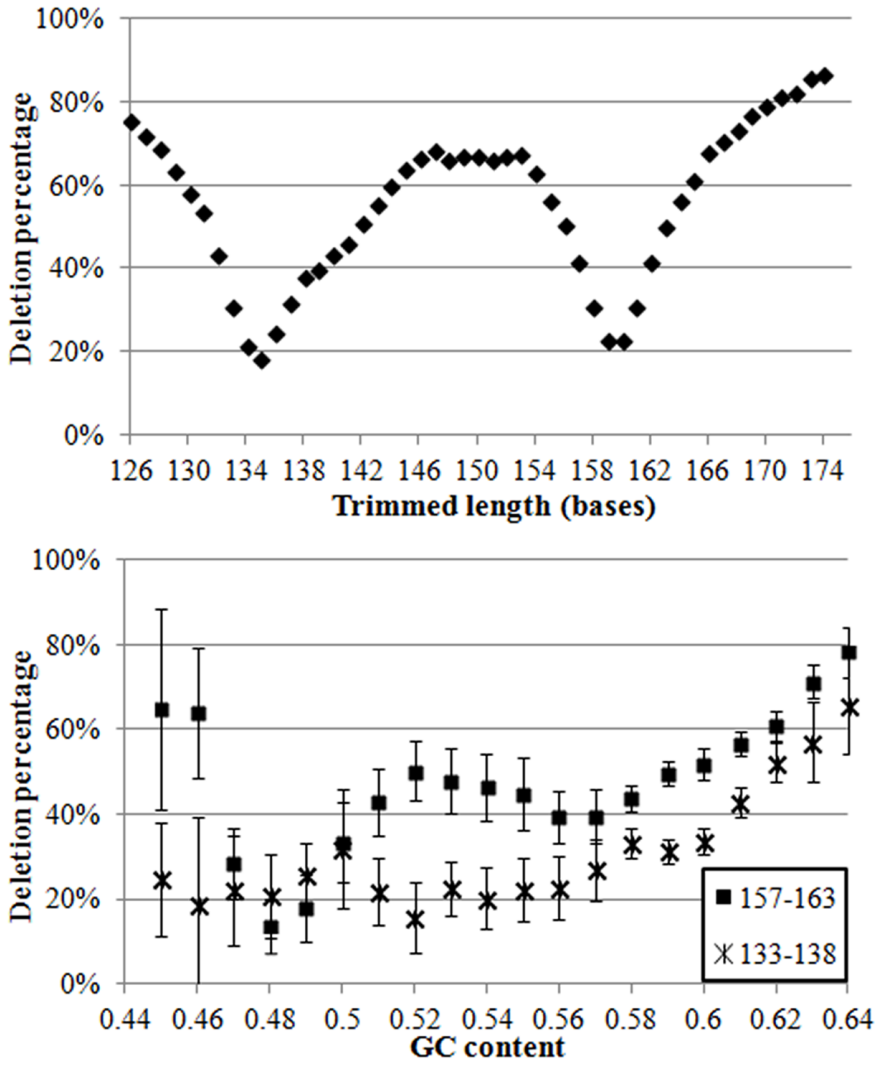

Figure 5-5 : Proportion of the sequences deleted when applying an average quality filter (q > 20) over the Ion Torrent dataset. Upper: Deletion percentages for the different trimmed lengths. Lower: Observing that the sequences with lengths ranging between 133-138 and 157-163 bases contained more than half of all the sequences and were of particularly good quality, these sequences were extracted and ordered by increasing GC content, individually for each length. Results, reported individually for the two clusters (133-138 and 157-163), correspond to the average deletion percentage for the different lengths in each cluster. Error bars correspond to the 95% confidence intervals calculated for all the lengths in each cluster.

The shape of the curve presented in the upper part of Figure 5-5 corresponds relatively well with the expected sequence lengths (major peaks at 136 and 161 bp and minor ones at 138, 141 and 156 bp). Many of the sequences outside these peaks can be hypothesized to contain insertions or

deletions (indels), thus generating variations in the sequence lengths. The shape of the curves presented on the bottom part of Figure 5-5 also supports this affirmation. Indeed, both curves presented a minimum deletion percentage for a certain GC content (at ~48% for the 157-163 bp cluster and between ~51-56% for 133-138 bp sequences), and these percentages increased when moving to the left or to the right on the GC content axis. These minimums probably corresponded to the OTUs' expected GC content, while the deletion increases can be explained by indels of AT or GC bases. These deletion increases were a lot more important for the longer sequences, indicating that length had a negative impact on the average quality of the sequences.

Interestingly, although length negatively affected the average quality of the sequences, the deletion percentage of the ~48% GC bin (157-160 bp) is still slightly lower than the percentages observed for the 51-56% GC bins (133-138 bp). These differences are, however, not significant. At GC contents above 58%, the relation between GC content and quality is significant for both length clusters. Also, the deletion percentages corresponding to the 157-163 bp sequences were significantly higher than those corresponding to shorter sequences for GC contents higher than 58%, confirming the premise that length decreases sequence quality. The same quality decrease seems to be observed for sequences having a GC percentage lower than 48%. However, since very few sequences corresponded to these GC bins, these deletion percentages should be considered with caution.

5.2.4.4 Comparing the outcomes of diversity studies based on DGGE and Ion Torrent sequencing

Until now, the values of the diversity indices produced by DGGE and Ion Torrent sequencing were quantitatively compared. While this methodology made it possible to identify potential biases in the results of both techniques, the main objective of this paper was to find out if the DGGE RADs elongation framework could yield quantitative conclusions similar to those found by next-generation sequencing technologies. Tableau 5-4 presents the linear correlation between Ion Torrent and DGGE diversity indices when applying various "correction" methodologies. These methodologies aimed at reducing the biases in the datasets, making the indices more comparable.

Tableau 5-4 : Slopes and coefficients of correlation of the linear regressions conducted between Ion Torrent sequencing and DGGE-based diversity estimates. All the coefficients were determined by forcing the intercept of the regression line to be the origin. For the Ion Torrent data, uncorrected refers to the diversity indices calculated directly from the clustered datasets, normalized corresponds to the values produced by sampling all the datasets 5,000 times and corrected values were produced with the DGGE RADs sampling framework. The numbers 2, 3 and 4 refer to the dissimilarity level (number of bases) used for the initial pseudo-single linkage algorithm (Huse et al., 2010). Except for the three lines in bold characters, all linear correlations were significant ($p < 0.05$).

Index	Ion Torrent	DGGE (ng well^{-1})	Slope	r
Simpson (1/D)	Uncorrected 3	400	1.92	0.92
	Uncorrected 3	200	1.43	0.94
	Normalized 3	200	1.51	0.94
	Corrected 4	200	1.93	0.96
	Corrected 3	200	1.37	0.95
	Corrected 2	200	0.86	0.91
Shannon ($exp^{H'}$)	Uncorrected 3	400	1.03	0.81
	Uncorrected 3	200	0.93	0.80
	Normalized 3	200	1.42	0.88
	Corrected 4	200	0.81	0.96
	Corrected 3	200	0.64	0.94
	Corrected 2	200	0.51	0.93
OTU richness	**Uncorrected 3**	**400**	**1.27**	**-**
	Uncorrected 3	**200**	**1.22**	**-**
	Normalized 3	**200**	**2.45**	**0.67**
	Corrected 4	200	0.68	0.91
	Corrected 3	200	0.62	0.90
	Corrected 2	200	0.59	0.90

If both techniques could produce accurate diversity estimates, the slope and the coefficient of correlation of the linear regression analysis would both be close to one. Considering that two very different techniques were compared and that biases were identified for both techniques, the slopes of the regression lines are rarely close to one (Tableau 5-4). However, the coefficients of correlation were still very high, indicating that a significant linear relation links the Ion Torrent and DGGE diversity indices. In order to find significant correlations for OTU richness (and, to a lesser extent, for $\exp^{H'}$), it is imperative to normalize the datasets to the same size. Interestingly, dataset normalization at 5,000 reads was less efficient than the correction framework. This can probably be explained by the nonlinear sample-dependent relationship between depth and richness (and $\exp^{H'}$). Even though certain parameters led to slopes close to one, this should be mostly seen as coincidence. Indeed, since no parameter yielded slopes approaching unity for 1/D at the same time, it seems obvious that the RADs produced by both techniques were never similar in terms of dominance profiles and lengths at the same time.

5.2.5 Discussion

5.2.5.1 Challenges associated with the use of DGGE as a quantitative tool

Based on *in silico* DGGE profiles produced from sequencing datasets of soil bacterial communities, it was demonstrated that the analytical methodology used in this paper unravelled DGGE banding patterns into relevant community dominance profiles (Lalande, Villemur, et al., 2013). Recognizing similarities in the sequencing-based RADs of many different soil samples, an elongation framework to accurately estimate diversity indices from DGGE profiles was developed. Of course, working on real soil samples is much more challenging, and certain biases linked to PCR amplification (Polz and Cavanaugh, 1998; Suzuki and Giovannoni, 1996; Takahiro, 2003) or the co-migration of DNA sequences on the DGGE profiles (Schmalenberger and Tebbe, 2003) can influence the outcome of DGGE-based diversity studies. However, biases linked to PCR amplification can be reduced by optimizing the PCR conditions and, even if co-migration was observed to occur on *in silico* DGGE profiles, it was still possible to find very strong correlations between DGGE and sequencing-based diversity estimates ($R^2 > 0.99$).

An element that could influence the outcome of DGGE-based diversity studies that, to the authors' knowledge, has never been described is the local saturation of DGGE gels. Figure 5-3 made it clear

that gel saturation has the potential to hinder the brightest DGGE bands from reaching their full intensity. As a dominance index, 1/D proved to be affected by gel saturation. Even though this process was visible when loading as little as ~75 ng of *E. coli* 16S rRNA gene PCR product, saturation could occur in certain DGGE profiles for much smaller peaks if they are located in high background noise intensity regions. Trying to determine whether gel saturation can be expected to be a widespread problem in previously published papers, a review was conducted to find out how much DNA researchers usually load in their DGGE wells. Surprisingly, this parameter is rarely reported and the few authors who did mentioned it loaded between 300 (Li et al., 2008) and 800 ng well^{-1} (Díez, Pedrós-Alió, Marsh, and Massana, 2001). The fact that saturation occurred for some samples when loading 400 ng well^{-1} leads to the hypothesis that the problem is probably quite common. As the results presented in this paper suggest, loading less DNA is a simple way to remedy the issue. If DGGE is to be used quantitatively, all gels should be run using a marker known to be below saturation. If less DNA is to be loaded on the gel, it can also become more important to color the DGGE gels with sensitive dyes (i.e. SYBR gold) to avoid limitations linked to fluorescence thresholds.

Another element that may influence the conclusions of DGGE-based diversity studies is primer quality and cycling conditions. Considering a coupling efficiency of 99% when producing oligonucleotides (Life Technologies, 2013), the synthesis of a 17-bases primer can yield over 15% erroneous/incomplete sequences. If annealed, the erroneous/incomplete primers could potentially blur the DGGE profiles and consistently produce multiple bands from a unique template. This will artificially inflate the apparent diversity of the studied communities and generate complex background noise profiles that are hard to delimit. If the elongation framework presented in Lalande et al. (2013) and used here to calculate the DGGE-based diversity indices is to be applied, it is imperative to be able to draw a realistic background noise profile in order to estimate accurate PSR values. The problem can be tackled by using purified primers and carefully optimized PCR conditions. Indeed, the HPLC purified primers and optimized PCR conditions used to produce the 200 ng well^{-1} gels led to simplified profiles (fewer weak bands).

5.2.5.2 Challenges associated with Ion Torrent sequencing

In this paper, the challenges associated with the use of the Ion Torrent PGM to study the diversity of soil bacterial communities were all related to the quality of the sequences. The shape of the

predicted per-base quality score produced from the current dataset was highly similar to what Loman et al. (2012) observed. However, in this case, the decrease in base quality was observed at a lower length than what was presented by the authors (at position ~100 instead of ~125). This 25-base difference seems to correspond to the length of the forward primer and barcodes used in this study.

The Ion Torrent PGM was reported to generate indel errors at a rate of 1.5 per 100 bases (Loman et al., 2012). Furthermore, Quail et al. (2012) reported a similar error rate of 1.78 per 100 bases. Considering that the average length of the sequences was higher in this paper than those reported in the two publications and that the average per-base quality was low for all the positions greater than 100, the error rate in the Ion Torrent dataset used in this publication is probably slightly higher than what was previously reported, though it cannot be quantified. To tackle sequencing error difficulties, a pseudo-single linkage clustering algorithm was applied with dissimilarity levels of 2, 3 and 4 (Huse et al., 2010). Depending on the sequence lengths (136 or 160 bases), the levels correspond to error rates of 1.3-1.5%, 1.9-2.2% and 2.5-3.0%, respectively. Whether or not these levels are suitable for this particular Ion Torrent dataset remains an open question. Since the datasets of the different samples were constructed with sequences of various lengths, it is unclear whether a single dissimilarity level can be seen as suitable for all the samples or even if a single level can be seen as suitable for a given sample. Considering the sensitivity of the diversity indices to sequencing errors (Tableau 5-3), it would be important to develop/adapt denoising tools for the newer Ion Torrent platform. Such tools exist for the 454 pyrosequencing platforms (i.e. AmpliconNoise (Quince et al., 2011)). Even if information on the topic is very scarce, it seems that the methods require adaptations before being applicable on Ion Torrent results. Generally speaking, assessing the accuracy of the diversity indices calculated from the sequencing datasets is not feasible. Indeed, sampling depth (Figure 5-4), sequencing errors (Tableau 5-3) and sequences length or GC content (Figure 5-5) were all proven to influence the diversity indices in a sample-dependent way. Depending on which bias is quantitatively the most significant for a given sample and index, the diversity of the sample can be underestimated or overestimated.

Finally, next-generation sequencing technologies have been reported to produce biased results, often in relation to the sequences' GC content (Jaenicke et al., 2011; Pinto and Raskin, 2012). These authors encouraged more research to determine whether the bias can cause important intra-sample distortions in the relative abundances of OTUs (Pinto and Raskin, 2012). Since the real diversity

of the studied samples is unknown, it is difficult to unambiguously state that GC content or, more generally, quality considerations have generated such distortions. However, many indicators point in this direction. The low deletion of sequences with GC content near 48% and, alternatively, the high deletion of all the sequences above a GC content of 58% is one such indicator (Figure 5-5). The fact that these low GC sequences are of slightly better quality than sequences with GC content between 51-56% even though they are nearly 30 bases longer is another (Figure 5-5). The fact that dataset processing removed proportionally fewer sequences for the samples dominated by low GC OTUs points in the same direction (Tableau 5-2). The strong linear correlation found between Ion Torrent and DGGE 1/D values (Tableau 5-4) cannot be interpreted as a lack of significant distortions. Indeed, the index was profoundly influenced by sequencing errors (Tableau 5-3), and it was hypothesized that sampling depth (through the occurrence of erroneous sequences) may influence 1/D to a greater degree than what Figure 5-4 suggests. It is also important to mention that, like sampling depth and sequencing errors, these distortions depend on the GC content and length of the OTUs in the dataset and are therefore sample-dependent.

Many aspects can be used to criticize the results produced by the Ion Torrent PGM. However, it must be acknowledged that the platform is newer than most of the other sequencing technologies and is still rapidly evolving. Since the time that *thesee* datasets were sequenced, a new emPCR template kit was released, potentially enabling the amplification of ~300 bases amplicons. In a press release, Life Technologies already announced that the read length capabilities of the PGM reached 400 bases with good quality scores throughout the sequences' lengths (Life Technologies, 2012).

5.2.5.3 Can DGGE and sequencing technologies yield similar ecological conclusions when studying complex soil bacterial communities?

While the accessibility and high-throughput of DGGE is hardly debatable, its ability to produce environmentally relevant diversity estimates for soil bacterial communities remains unclear. Focusing only on the communities' dominant phylotypes, traditional fingerprinting techniques are known to produce diversity estimates that are not correlated to the diversity of the complete communities (Blackwood et al., 2007; Lalande, Villemur, et al., 2013). In the authors' opinion, if only dominants are to be considered, DGGE should not be used to study the diversity of soil bacterial communities.

The story is different when using the elongation framework presented in this paper. Considered to be the best available technology, DNA sequencing was used with the aim of validating the framework. In doing so, biases linked to DGGE profiling and especially to the Ion Torrent datasets were clearly identified. Since the *true* diversity of the studied samples is unknown, it is impossible to determine which technology produces the most accurate diversity estimates. While divergences exist, it is important to note that the diversity estimates produced by both technologies are highly correlated when the bias linked to sampling depth is corrected (Tableau 5-4). Considering the variability of OTU richness with respect to the correction methodology, it is not advisable to recommend this indicator to compare the diversity of soil bacterial communities. This supports a previous conclusion stating that the actual technologies do not allow the production of richness-based metrics suitable to be used to quantitatively compare the diversity of soil bacterial communities (Pinto and Raskin, 2012).

Unlike richness, $\exp^{H'}$ and especially 1/D estimates were very robust. Even before bias correction, the DGGE and Ion Torrent sequencing indices were significantly correlated ($p < 0.05$). However, these high correlations do not mean that the indices were accurate. Questions can be raised about the accuracy of the Ion Torrent diversity estimates. Indeed, even though it is hard to tell which of LM_{cln} or S_{cln} is the most diversified (in terms of dominance) when comparing the DGGE profiles of both samples, the Ion Torrent 1/D value calculated for sample LM_{cln} is more than 30% higher than the value corresponding to S_{cln} (Tableau A4-2). In contrast, DGGE ranked these samples as almost equally diversified. The same was observed for LM_{3000} and LM_{300}; the sequencing-based 1/D corresponding to LM_{300} was almost two times lower than the value corresponding to LM_{3000}.

In conclusion, contrary to what the discussion may seem to suggest, next-generation sequencing technologies are undoubtedly much more powerful than DGGE. Indeed, this paper only sought to carry out a comparative assessment of the diversity of soil bacterial communities. If phylogenetic information is also required, DGGE is completely outcompeted by sequencing technologies. Results have shown that the Ion Torrent platform (and maybe other platforms, too) must still account for potential distortions and other quality considerations. Of course, the magnitude of the DGGE-based diversity estimates can also be seen as uncertain. However, these values were found to be highly repeatable. Overall, both approaches would have led to the same conclusions if the impact of PCP contamination was quantified. This is significant considering that the DGGE RADs elongation framework was developed from *in silico* profiles drawn using GS FLX datasets, while

this paper analyzed some very challenging experimental DGGE profiles and compared the results with diversity estimates drawn from a dataset produced by an Ion Torrent PGM. The results are promising for researchers using DGGE and those using next-generation sequencing technologies. Even though the two methodologies are quite different, they lead to similar portraits of soil bacterial communities.

5.2.6 **Acknowledgements**

The authors would like to sincerely thank Sylvie Sanschagrin and Christine Maynard for their support in producing the sequencing datasets. The authors acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada. The CIRAIG would also like to thank its industrial partners for their financial support: ArcelorMittal, Bombardier, Bell Canada, Cascades, Eco Entreprises Québec, RECYC-QUÉBEC, Groupe EDF, Gaz de France, Hydro-Québec, Johnson & Johnson, LVMH, Michelin, Mouvement des caisses Desjardins, Nestlé, Rio Tinto Alcan, RONA, SAQ, Solvay, Total, Umicore, and Veolia Environment.

CHAPITRE 6 DÉTERMINATION DE L'EFFET ÉCOTOXIQUE DU PENTACHLOROPHÉNOL SUR LA DIVERSITÉ DES COMMUNAUTÉS BACTÉRIENNES DES SOLS

Cette section est centrée autour du troisième manuscrit intitulé « *Assessing the impact of pentachlorophenol on soil bacterial communities using ecological diversity indices* » (section 6.2) soumis à la revue *The ISME Journal* (soumis le 8 mai 2013).

6.1 Lien entre le deuxième et le troisième manuscrit

Le second manuscrit a permis d'établir qu'une corrélation linéaire significative existe entre les estimés de diversité produits à partir du DGGE et de l'Ion Torrent pour la richesse ainsi que pour les indices de Shannon ($exp^{H'}$) et de Simpson (1/D). Puisque la diversité réelle des communautés étudiées dans le second manuscrit n'était pas connue, il n'est pas possible d'affirmer que l'une ou l'autre des approches a la capacité de produire des estimés de diversité quantitativement représentatifs des communautés étudiées. Par contre, la détermination de l'effet écotoxique du PCP requiert de comparer la diversité de deux échantillons (sol propre vs sol contaminé). Dans cette optique, puisque les indices de diversité des deux approches étaient significativement corrélés, l'utilisation de l'une ou l'autre des techniques mènerait à des conclusions écologiques similaires.

Dès lors, la méthode d'élongation des RAD DGGE a été utilisée afin de déterminer l'effet du PCP sur les communautés bactériennes des sols. Afin de pouvoir comparer la réponse obtenue à partir des indices de diversité à celle générée par l'utilisation d'un indicateur fonctionnel, le potentiel d'activité hydrolytique de la FDA des échantillons de sol a également été déterminé. Cet indicateur a été choisi puisqu'il se mesure rapidement, facilement et qu'il est corrélé aux mesures de biomasse microbienne, un indicateur s'étant montré sensible au PCP. Ce manuscrit a donc cherché à comparer la sensibilité des mesures basées sur la diversité des communautés bactériennes à celle du potentiel d'activité hydrolytique de la FDA. Aussi, vu la très grande diversité génétique des communautés bactériennes des sols, une courbe concentration-réponse tracée à partir d'une mesure de diversité bactérienne peut s'apparenter aux courbes SSD tracées afin d'estimer la sensibilité des écosystèmes face aux contaminants. Le manuscrit vise donc également à comparer la sensibilité des communautés biotiques terrestres à l'échelle microscopique et macroscopique.

6.2 "Assessing the impact of pentachlorophenol on soil bacterial communities using ecological diversity indices"

6.2.1 Abstract

Species sensitivity distribution (SSD) is a central concept in assessing the impact of xenobiotics on natural ecosystems and deriving environmental quality criteria. Data scarcity is a major limitation to SSD modeling, especially for terrestrial ecosystems. Until now, the impact of pentachlorophenol (PCP) on soil bacterial communities was mostly evaluated using functional indicators, yielding highly variable results. The aim of this paper was to assess the usefulness of ecological diversity indices to model the impact of pentachlorophenol on soil bacterial communities. Two different soils were contaminated with PCP at seven different concentrations. The effect of PCP on soil bacterial diversity was measured using a new framework that made it possible to estimate soil bacterial diversity using PCR-denaturing gradient gel electrophoresis (DGGE) technology. Three diversity-based indicators were compared to the assay of fluorescein diacetate (FDA) hydrolytic activity—a simple measure well correlated with microbial biomass. Results showed that the Shannon ($\exp^{H'}$) and Simpson (1/D) indices were at least 25 times more sensitive to PCP than the FDA assay. Furthermore, it was demonstrated that the $\exp^{H'}$ and 1/D concentration-response curves where highly similar to the SSD corresponding to PCP in terrestrial ecosystems and that the diversity-based 10% effect concentrations (EC_{10}) were comparable to PCP soil screening levels adopted by various regulatory agencies. If these observations may be generalized to other contaminants, this cost-effective methodology has the potential to help fill many gaps in terrestrial ecotoxicological databases and, more broadly, in the understanding of the impact of xenobiotics (or complex mixtures) on terrestrial ecosystems.

6.2.2 Introduction

A species sensitivity distribution (SSD) is a statistical distribution, fitted from ecotoxicological data on a limited set of species, used to infer the impact that a contaminant could have on an entire ecosystem (Posthuma, Traas, and Suter II, 2001). Policymakers use the framework to derive environmental quality criteria (EQC). For acceptable results, SSD modeling requires that a sufficient number of species be tested for a given contaminant (Dowse, Tang, Palmer, and Kefford, 2013). The minimum size of the datasets was set at 10 data points (Wheeler, Grist, et al., 2002) and

should ideally cover as many taxonomic groups as possible (von der Ohe and Liess, 2004). Due to data scarcity, the number of available data is generally much lower than 10 when focusing on terrestrial ecosystems, and most of the data pertains to standard test species (Frampton et al., 2006). The derivation of relevant EQCs requires the production of more data on terrestrial taxonomic groups not included in the databases.

Soil microbial communities are among the most diverse and abundant on Earth (Gans et al., 2005) and play key roles in terrestrial ecosystems (van der Heijden et al., 2008). Though numerous papers have studied the impact of xenobiotics on these communities, they are hardly represented in ecotoxicological databases (Suter II et al., 2001). This lack of data can probably be linked to technical difficulties in studying such complex communities and small organisms. Earlier papers focused on functional parameters in order to evaluate the adverse effects of xenobiotics (Filip, 2002). However, these indicators present several limitations when used as toxicity indicators. Indeed, some indicators increased following contaminants addition to soils (Macur et al., 2007; Scelza et al., 2008). Other parameters, such as enzyme activity, were directly inhibited by certain substances, making it difficult to differentiate between toxicity and inhibition (Vig, Megharaj, Sethunathan, and Naidu, 2003).

The advent of molecular microbiology opened the way, through many different fingerprinting techniques (Imfeld and Vuilleumier, 2012; Kirk et al., 2004), to the study of soil bacterial community diversity. Focusing solely on the most abundant phylotypes, the methods do not produce estimates correlated to the actual diversity of the communities (Blackwood et al., 2007; Lalande, Villemur, et al., 2013). Even though next-generation sequencing technologies make it possible to study microbial communities with unprecedented depth (Weinstock, 2012), the costs, sequencer availability and potential biases in sequencing datasets (Jaenicke et al., 2011; Nacke et al., 2011; Pinto and Raskin, 2012) can still be problematic. To overcome these limitations, a new framework to accurately quantify soil bacterial community diversity using a PCR-denaturing gradient gel electrophoresis (DGGE) methodology—a widespread, low-cost and high throughput fingerprinting technique—was recently published. The methodology was developed on *in silico* DGGE migration profiles synthesized using pyrosequencing datasets (Lalande, Villemur, et al., 2013) and then further validated on real samples (Lalande, Yergeau, Greer, Villemur, and Deschenes, 2013). The framework yielded diversity estimates highly representative of the studied bacterial communities.

While some papers have studied the impact of pentachlorophenol (PCP) on soil bacterial communities, most targeted integrative microbe-mediated functions with variable results (Martí et al., 2011; Megharaj et al., 1998; Scelza et al., 2008). Since soil bacterial communities are composed of thousands of interacting species all different in terms of their ecological niche and metabolic potential (Stres and Tiedje, 2006), assessing the impact of PCP using ecological diversity indices (Magurran, 2004) shows great potential. By fractioning these communities into species instead of considering them as a whole, diversity indices can potentially be more sensitive than functional measures. Furthermore, if the range of sensitivity of bacterial species to PCP is similar to what is observed for higher species, a single test could yield enough information to derive relevant EQCs.

The aim of this paper was to assess the usefulness of indicators based on the diversity of soil bacterial communities to assess the toxicity of PCP. The contaminant was mainly chosen because many authors have studied its toxicity to microbial communities and other terrestrial species. Through concentration-response (C-R) modeling, the sensitivity of diversity indices in detecting PCP was compared to an integrative functional indicator: the soils' fluorescein diacetate (FDA) hydrolytic activity. The FDA hydrolysis assay was chosen because it correlates well with microbial biomass (Swisher and Carroll, 1980) and is widely accepted as a simple method to measure total microbial activity (Adam and Duncan, 2001). Furthermore, by comparing the diversity-based C-R curves to the terrestrial SSD drawn for PCP, the ability of such indicators to derive relevant EQCs was assessed.

6.2.3 Material and methods

6.2.3.1 PCP-contaminated soil microcosms

Approximately 2 kg of sandy loam (LM) and sandy (S) soils were sampled near Montréal (Québec, Canada), immediately sieved (2 mm mesh) and stored at 4°C until use. The particle size distribution was determined according to the D1140-00 (ASTM International, 2006) and D422-63 (ASTM International, 2007a) ASTM methods. The soils were classified according to the USDE classification system (sand [2mm-50µm], silt [50-2µm], and clay [<2 µm]). Total carbon (TC) was measured through combustion with an induction furnace (LECO Corporation, St-Joseph, MI) according to the D2974-07a (ASTM International, 2007b) ASTM method. Total inorganic carbon (TIC) was performed with a phosphoric acid treatment followed by an infrared determination. Total

organic carbon (TOC) was calculated from the difference between TC and TIC. Soil pH was determined with distilled water using potentiometry (Orion Surfow semi-micro Ross combination pH electrode) according to the D4972-01 (ASTM International, 2007c) ASTM method with a soil to water ratio of 1:2. Soil granulometry, total organic carbon content and pH are presented in Table 1.

For each soil, twenty-one soil microcosms were derived, representing seven different levels of PCP contamination (all prepared in independent triplicates). Three other microcosms served as control and were not spiked with PCP. Based on the DGGE profiles published by Martì et al. (2011), contamination was performed by dissolving 0, 0.6, 1.3, 2.6, 6.3, 12.3, 25.9 and 89 mg of PCP (Sigma-Aldrich, purity > 97%) in 5 mL of dichloromethane (HPLC grade) in empty 250 mL amber bottles. The bottles were gently shaken during dichloromethane volatilization in order to recover the PCP on the bottle walls. After dichloromethane volatilization, the soils (~ 30 g dry wt) were added to the bottles and mixed in a tumble action agitator for 24 hours in order to recover PCP from the bottle walls. Samples were humidified with distilled water to 60% (w/w) of the soil water holding capacity and manually mixed. Soils were kept at room temperature and protected from light for 28 days prior to DNA extraction and FDA hydrolytic activity assessment.

Tableau 6-1 : Physico-chemical properties of soils.

Soil	% sand	% loam	% clay	Total organic carbon (%)	pH
LM	58	24	18	2.8 ± 0.4	7.39 ± 0.03
S	98	2	0	2.2 ± 0.1	6.68 ± 0.03

The PCP concentrations at the beginning of the incubation period were determined (Table 2) as previously described (Pu and Cutright, 2006). Since ecotoxicological data are usually determined from the initial concentrations, the concentrations after 28 days were not measured. Soils (10 g) were dehydrated with 5 g of anhydrous magnesium sulfate. A surrogate consisting of 100 µL of a solution of 2,4,6-tribromophenol (1 g/L) was added into each sample. PCP was extracted from the soils with a Tecator Soxtec System HT 1043 extraction unit (Foss NA, Eden Prairie, MN, USA) using dichloromethane (HPLC grade) as a solvent (10:1 ratio (v/w)). After dichloromethane volatilization, the PCP was solubilised in hexane (HPLC grade) and measured using a gas chromatograph equipped with a flame ionisation detector (GC-FID) (Agilent model 6890N, USA).

The GC was equipped with a 60 m X 0.25 mm FactourFour capillary column (Varian) with a 0.25 µm stationary phase. The mobile phase was helium, and the flow rate was set at 29 mL/min. The injection was made at an initial temperature of 69°C for 0.2 min and was increased at a rate of 200°C/min until 300°C. The temperature of the detector was 320°C. The initial temperature of the oven was 30°C and increased at a rate of 10°C/min until 300°C. That temperature was maintained for 10 min. The concentrations in the samples corresponding to levels 0, 1, 2 and 3 were below the detection limit of the quantification method (~ 100 mg/kg), and the samples were sent to an independent laboratory (Maxxam Analytics, Saint-Laurent, Québec, Canada). PCP was quantified using a gas chromatograph coupled with a mass spectrometer (GC-MS - detection limit ~ 0.1 mg/kg). The measurements were taken on composite samples prepared for each soil and contamination level.

Tableau 6-2 : Initial PCP concentrations measured in the soil samples. Results are presented as the mean PCP concentration for each level, with the associated 95% confidence intervals (calculated whenever possible).

Sample	PCP concentration (mg PCP/kg dry wt soil)	Sample	PCP concentration (mg PCP/kg dry wt soil)
LM_0	< 0.1	S_0	< 0.1
LM_1	10	S_1	7
LM_2	24	S_2	28
LM_3	44	S_3	44
LM_4	131 ± 1	S_4	129 ± 7
LM_5	182 ± 4	S_5	194 ± 19
LM_6	420 ± 19	S_6	440 ± 29
LM_7	1774 ± 35	S_7	2111 ± 59

6.2.3.2 Quantification of the diversity of the soil bacterial communities

DNA extractions were carried out as previously described (Bürgmann et al., 2001), with minor modifications. Soil (300 mg) and 0.1 mm diameter zirconia-silica beads were mixed in a 1:1 (w/w)

ratio. Bead-beating was done at 2 800 rpm for 90 seconds with a Mini BeadBeater-8 (BioSpec Products Inc, Bartlesville, OK, USA) in 650 µl of extraction buffer composed of 0.2% (w/v) hexadecyltrimethylammonium bromide (CTAB), 1 mM dithioteitrol, 0.2 M sodium phosphate buffer (pH 8), 0.1 M NaCl and 50 mM ethylenediaminetetraacetic acid. Purification steps involved polyvinylpolypyrrolidone (PVPP) spin columns, incubation/centrifugation in a 20% (w/v) polyethylene glycol and 2.5 M NaCl solution and phenol:chloroform:isoamyl alcohol extraction (Thakuria et al., 2008). DNA was recovered by ethanol precipitation (Sambrook et al., 1989). Purified DNA was quality-checked using a 1% (w/v) agarose gel and quantified with a TECAN GENios fluorometer (Tecan Group Ltd., Männedorf, Switzerland). Samples were stored at -20°C.

PCR amplifications (V3 region) of the bacterial 16S rRNA gene were done using primers 341f (5'-CCTACGGGAGGCAGCAG-3') and 534r (5'-ATTACCGCGGCTGCTGG-3') (Muyzer et al., 1993). Both primers were HPLC purified. The forward primer also contained at its 5' end a 40 nt GC clamp (5'-CGCCCGCCGCGCGCGGCGGGCGGGGCGGGGGCACGGGGGG-3') (Muyzer et al., 1993). Amplifications were conducted in 50 uL volumes using the 2X Phusion® High-Fidelity PCR Master Mix (Fisher Scientific Inc.). PCR mixtures contained 200 µM of each dNTP, 2.5 mM of MgCl2, 0.6 µg/µL of bovine serum albumin, 0.1 µM of each primer and 5 ng of extracted DNA. Amplification was carried out with a Techne Genius Thermocycler (Bibby Scientific Ltd., Staffordshire, UK). After an initial two-minute denaturation step at 94°C, a total of 35 amplification cycles were carried out, consisting of 10 second denaturation at 94°C, 10 second annealing and 10 second elongation at 72°C. The amplification process was concluded by a final elongation step of 5 minutes at 72°C. To increase the specificity and yield of the amplification process (Korbie and Mattick, 2008), the annealing temperature was decreased from 67.5 to 62.5°C during the first 10 cycles and then kept at 62.5°C for the remaining 25 cycles.

DGGE profiling was conducted using a D-Code system (Bio-Rad Laboratories Inc., Hercules, CA, USA). The 8% (w/v) acrylamide gel was cast into 16 x 16 cm plates with 1 mm thick spacers and composed of a 32.5-72.5% denaturing gradient, 100% corresponding to 7 M urea and 10% (v/v) deionized formamide. Approximately 200 ng of amplified DNA was loaded onto the gel for each sample. A blank PCR sample and 60 ng of PCR product amplified from *E. coli* genomic DNA were also added in the first two wells of each gel. Electrophoresis was run at 60°C and 60 V for 16 hours. The gel was stained for 30 minutes in a fresh solution of SYBR Gold nucleic acid gel stain (Invitrogen Ltd) and photographed using a Quantum ST4 gel documentation system (Vilber

Lourmat, Marne La Vallée, France). The *E. coli* PCR product was used as a marker to ensure that gel saturation did not occur (Lalande, Yergeau, et al., 2013).

The diversity of soil bacterial communities was determined from the DGGE profiles using a framework that was previously described in detail (Lalande, Villemur, et al., 2013) and further validated using next-generation sequencing technologies (Lalande, Yergeau, et al., 2013). From the DGGE bands' relative intensities and the peak-to-signal ratio—a parameter extracted from the DGGE profile—the framework produced a rank-abundance distribution (RAD) representative of the studied soil bacterial community. This RAD contains information on the abundance of the different bacterial species that make up the community (see the initial publication for a complete description of the framework). As recommended by Jost (2006), bacterial community diversity was quantified using indices corresponding to an effective number of species (Shannon expH' and Simpson's 1/D and community richness). Community richness corresponded to the length of the RADs produced by the framework mentioned above. Using the abundance values of the RADs, the Shannon and Simpson indices were calculated using Eq.1 and Eq.2, respectively.

$$exp^{H'} = e^{-\sum_{i=1}^{S} p_i * \ln(p_i)} \quad [1]$$

$$\frac{1}{D} = \frac{1}{\sum_{i=1}^{S} p_i^2} \quad [2]$$

Where p_i corresponds to the relative abundance of the i^{th} species and S corresponds to community richness.

6.2.3.3 Assay for FDA hydrolytic activity

The assay for FDA hydrolytic activity was carried out as described by Green et al. (2006) with minor modifications. Briefly, 1.0 g of soil, 50 ml of 60 mM sodium phosphate buffer (pH 7.6) and 0.50 mL of 4.9 mM FDA lipase (Sigma-Aldrich) substrate solution were mixed in 125 mL Erlenmeyer flasks and placed in an incubator for 2 h at 37°C. After the incubation period, 2.0 mL of acetone was added to terminate the FDA hydrolysis. Then, using a 5 mL syringe, approximately 2.0 mL of the aqueous phase was filtered through a 0.45 µm PVDF membrane (EMD Millipore Corporation, Billerica, MA) directly into a colorimeter tube. Filtrate absorbance was measured on a Varian Cary Bio-100 spectrophotometer (Agilent Technologies Inc., Santa Clara, CA) at a

wavelength of 490 nm. For all samples, the measurements were taken in triplicate. Blank samples were also prepared, and their filtrate absorbance was subtracted from the samples' activity.

6.2.3.4 Concentration-response and SSD modelling

C-R modeling was undertaken using the extension package drc (Ritz and Streibig, 2005) developed for the open source statistical software R (R Developement Core Team, 2013). As recommended by Ritz (2010), no data preprocessing or normalization was done. Curve fitting was carried out using the Weibull (type 1 and 2), log-logistic and log-normal models (with two parameters). For each dataset, the model that produced the lowest residual standard error was used to draw the C-R curve.

For the SSD modeling, ecotoxicological data on the toxicity of PCP to terrestrial organisms were retrieved from the ECOTOX database (U.S. Environmental Protection Agency, 2013) and the TerraSys risk assessment software (Sanexen Services Environnementaux Inc., 2002). Considering the arguments of Warne and van Dam (2008), the no- or low-observed effect concentrations (NOECs and LOECs) were not used. In light of availability issues, the selected data were restricted to the effective or lethal concentrations to 50% of the tested populations (EC_{50} and LC_{50}, hereafter, the term EC_{50} will include both type of indicators). To include as many data as possible, the EC_{50} corresponding to all endpoints (emergence, growth, reproduction, mortality, etc.) were included in the dataset. The geometric mean was calculated if more than one data was available for a given test species. Considering that the different taxa represented in the dataset were unequally represented (dominated by worms and terrestrial plants), the SSD curve was drawn using the geometric mean of the ecotoxicological data available for the different taxonomic groups (Larsen and Hauschild, 2007). The dataset was then sorted by ascending geometric mean value, and a plotting position was assigned to each value using the Hazen formulae (Eq.3) (Fuglem, Parr, and Jordaan, 2013). The plotting position associates each ecotoxicological indicator with a cumulative probability value (%) on the *y-axis*. Since the dataset is considered to be representative of the entire terrestrial biological community, these cumulative probabilities can be seen as the cumulative percentage of species affected above their EC_{50} when exposed at a given concentration of PCP (value on the *x-axis*). This measure of risk is termed the *potentially affected fraction of species* (PAF).

$$p_i = 100 * \left(\frac{i-0.5}{N}\right) \quad [3]$$

Where p_i corresponds to the plotting position (in %) assigned to each ecotoxicological data, i corresponds to the data rank when the dataset is sorted in ascending order (i = 1, 2…N), and N corresponds to the size of the dataset (number of ecotoxicological data). Data fitting was done as described above. To quantitatively compare the different indicators, the PCP concentrations corresponding to distribution percentiles of 5, 10 and 50 (EC_5, EC_{10} and EC_{50}) were extracted from the C-R and SSD curves. The 95% confidence intervals associated with these percentiles were approximated using the delta method, which yielded accurate results in similar situations (Demidenko, Williams, Flood, and Swartz, 2012).

6.2.4 Results

6.2.4.1 Using diversiy indices to model the impact of PCP towards bacterial communities

Two soils were contaminated with increasing concentrations of PCP (Tableau 6-2) and incubated for 28 days. The two soils were chosen to assess the influence of physico-chemical properties on the impact of PCP on soil bacterial communities. Indeed, the bioavailability and, as such, the toxicity of PCP were proven to be influenced by the pH (Arcand et al., 1995) and composition (Martí et al., 2011; Puglisi et al., 2009) of the soils. After the incubation period, the soil microcosms were sampled to measure the FDA hydrolytic activity and extract total DNA. From this DNA, the bacterial diversity profiles were derived by PCR-DGGE (Figure 6-1). Even if they were produced from independently contaminated samples, the DGGE migration profiles were highly replicable (not shown). As PCP concentrations increased, bright bands—probably corresponding with bacteria populations able to tolerate or degrade PCP—appeared in the profiles. Interestingly, the LM_7 and, particularly, S_7 samples were not characterized by the same bright bands as the LM_6 and S_6 samples. This observation corresponds to a high PCP toxic effect, not show by the DGGE profile. The diversity increases observed for sample S_7 and, to a lesser extent, LM_7 (when compared with S_6 and LM_6) were associated with concentrations sufficiently high to affect the species generally able to degrade or tolerate PCP. The diversity indices calculated for these samples were therefore excluded from the data fitting process.

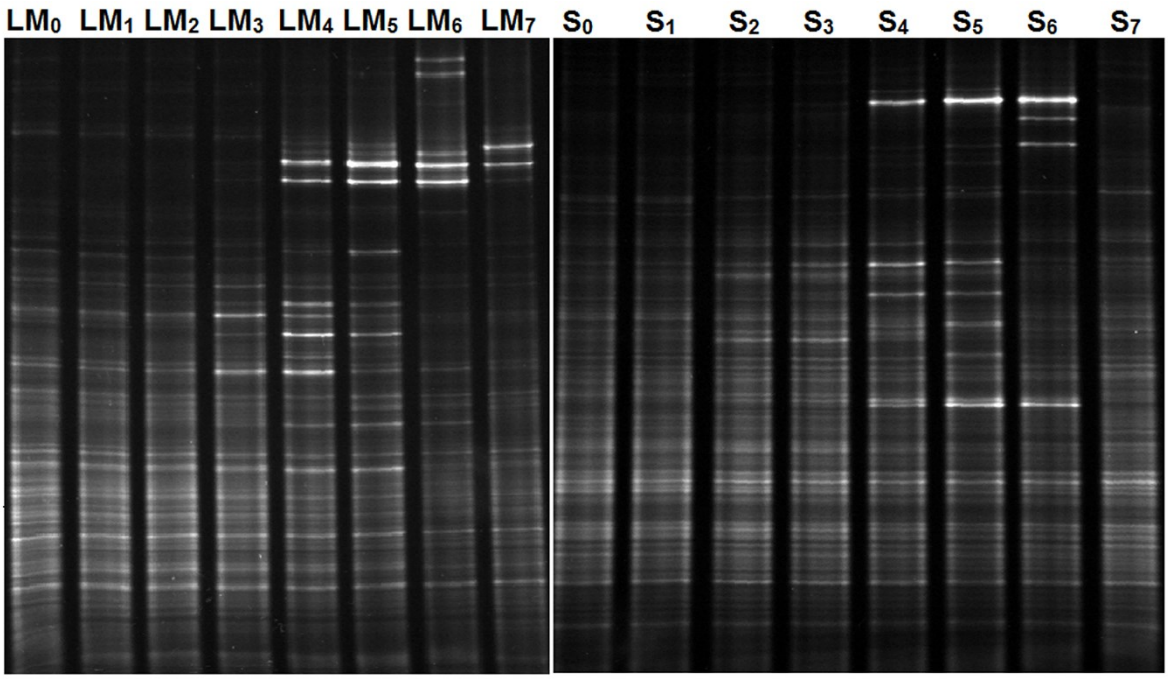

Figure 6-1 : DGGE migration profiles corresponding to the first replicate. DNA samples extracted from soil LM and soil S were subjected to PCR for the amplification of 16S rRNA gene sequences (the V3 variable region). The amplicons were separated on 8% polyacrylamide DGGE with a 32.5-72.5% denaturant gradient. The indices number in the sample names refer to the initial PCP concentrations in Tableau 6-2.

The DGGE migration profiles were used to measure three bacterial diversity indices: Simpson's 1/D, Shannon $\exp^{H'}$ and community richness. The impact of PCP can be seen in the reduction of these indices (Figure 6-2A, B and C). A reduction in the FDA hydrolytic activity was also noticed (Figure 6-2D). The four indices exhibited a typical sigmoidal response with increasing concentrations of PCP, from which the EC_{50} values (the PCP concentrations causing a 50% decrease of diversity or activity) were extracted (Tableau 6-3).

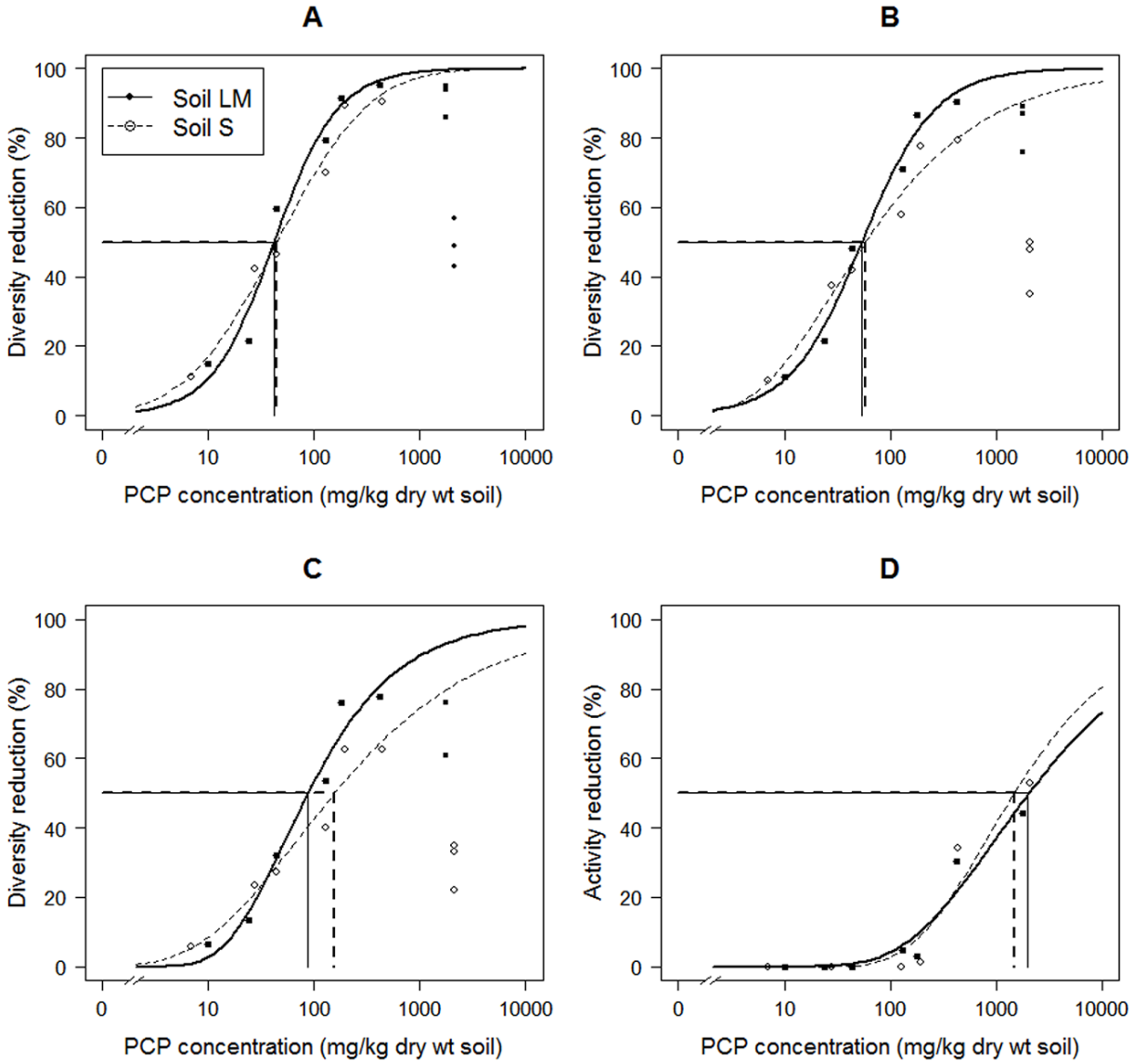

Figure 6-2 : Concentration-response (C-R) curves produced by the four indicators. The reduction percentages were calculated from the diversity or activity of the uncontaminated samples corresponding to a 0% reduction. **A.** Simpson's 1/D **B.** Shannon $\exp^{H'}$ **C.** Community richness **D.** FDA hydrolytic activity. For C-R modeling, the diversity indices calculated for the most contaminated samples were excluded from the data fitting process, since these diversity increases were associated with an important PCP toxicity not shown by DGGE profiles. The horizontal and vertical solid (soil LM) and dotted (soil S) lines show the position of the EC_{50} on the x-axis.

As demonstrated by the EC_{50}, the sensitivity of the different indicators was highly variable. The FDA hydrolytic activity was, by far, the least sensitive indicator of the impact of PCP

contamination. It is interesting to note that, except for community richness, both soils were equally sensitive to PCP.

Tableau 6-3 : EC_{50} values extracted from the C-R curves presented in Figure 6-2. The numbers in brackets correspond to the 95% confidence intervals.

Indicator	EC_{50} (mg PCP/kg dry wt soil)	
	Soil LM	Soil S
Simpson's 1/D	42 [35 ; 49]	45 [38 ; 52]
Shannon $exp^{H'}$	45 [38 ; 52]	59 [47 ; 70]
Community richness	88 [74 ; 102]	156 [113 ; 198]
FDA hydrolytic activity	1990 [1110 ; 2870]	1480 [990 ; 1970]

6.2.4.2 Using the diversity of soil bacterial communities to derive EQCs

For PCP in soils, ecotoxicological databases (Sanexen Services Environnementaux Inc., 2002; U.S. Environmental Protection Agency, 2013) contained ~80 different EC_{50} on 24 different terrestrial species. However, many of these data pertained to the same species. Furthermore, 10 of these EC_{50} were not extracted from fitted distributions but reported as the tested concentration that produced the impact closest to 50%. These values were deemed inaccurate and excluded from the dataset. As shown in Tableau 6-4, the final dataset contained toxicological information on some 14 different terrestrial species and was dominated by worms and terrestrial plants. The plant and nematode species were particularly sensitive to PCP. To avoid the overrepresentation of certain taxonomic groups in the SSD, the modeling process was carried out using the geometric mean of the EC_{50} available for each taxonomic group (*x-axis*) and the associated plotting positions (*y-axis*) (Tableau 6-4).

Tableau 6-4 : Ecotoxicological data used for SSD modeling. Considering that the dataset was dominated by worms and terrestrial plants, the SSD modeling process was carried out using the geometric mean of the EC_{50} available for each taxonomic group. All the ecotoxicological indicators (geometric mean and EC_{50}) are given in mg PCP/kg soil. The PAF (potentially affected fraction of species) corresponds to the cumulative proportion of species affected above their EC_{50} at a given PCP concentration (See the Material and methods section for the calculations).

Taxonomic group	Geometric mean of the EC_{50}	PAF (%)	Species name	EC_{50}	Number of data	Standard deviation
Terrestrial plants	22.6	12.5	*Lactuca sativa*	8.3	9	7.7
			Brassica rapa	9.9	2	1.9
			Raphanus sativa	55.7	4	30.1
			Avena sativa	57.3	3	75.3
Nematodes	23.1	37.5	*Panagrellus redivivus*	11.1	4	5.7
			Plectus communis	47.9	1	-
Arthropods	87	62.5	*Folsomia candida*	87	1	-
Worms	135.1	87.5	*Eisenia fetida andrei*	62.8	5	47.6
			Eisenia andrei	63.7	11	137.3
			Eisenia fetida	77.4	19	81.3
			Lumbricus terrestris	123.7	3	45.6
			Enchytraeus albidus	136	1	-
			Eudrilus eugeniae	158.5	1	-
			Lumbricus rubellus	996.9	6	637.9

The SSD curve describing the sensitivity of terrestrial species to PCP was compared to the C-R curves representing the effect of PCP on soil bacterial community diversity (1/D and $\exp^{H'}$ indices)

(Figure 6-3). The curves drawn using community richness were very different from the SSD and were therefore not presented. The SSD curve and the C-R curves corresponding to 1/D (both soils) and $exp^{H'}$ (soil LM) are highly similar. The C-R curve corresponding to soil S $exp^{H'}$ diverged from the SSD for percentiles higher than 50. When considering the 95% confidence intervals, the five curves presented in Figure 6-3 cannot be differentiated for all the percentiles between 5 and 50.

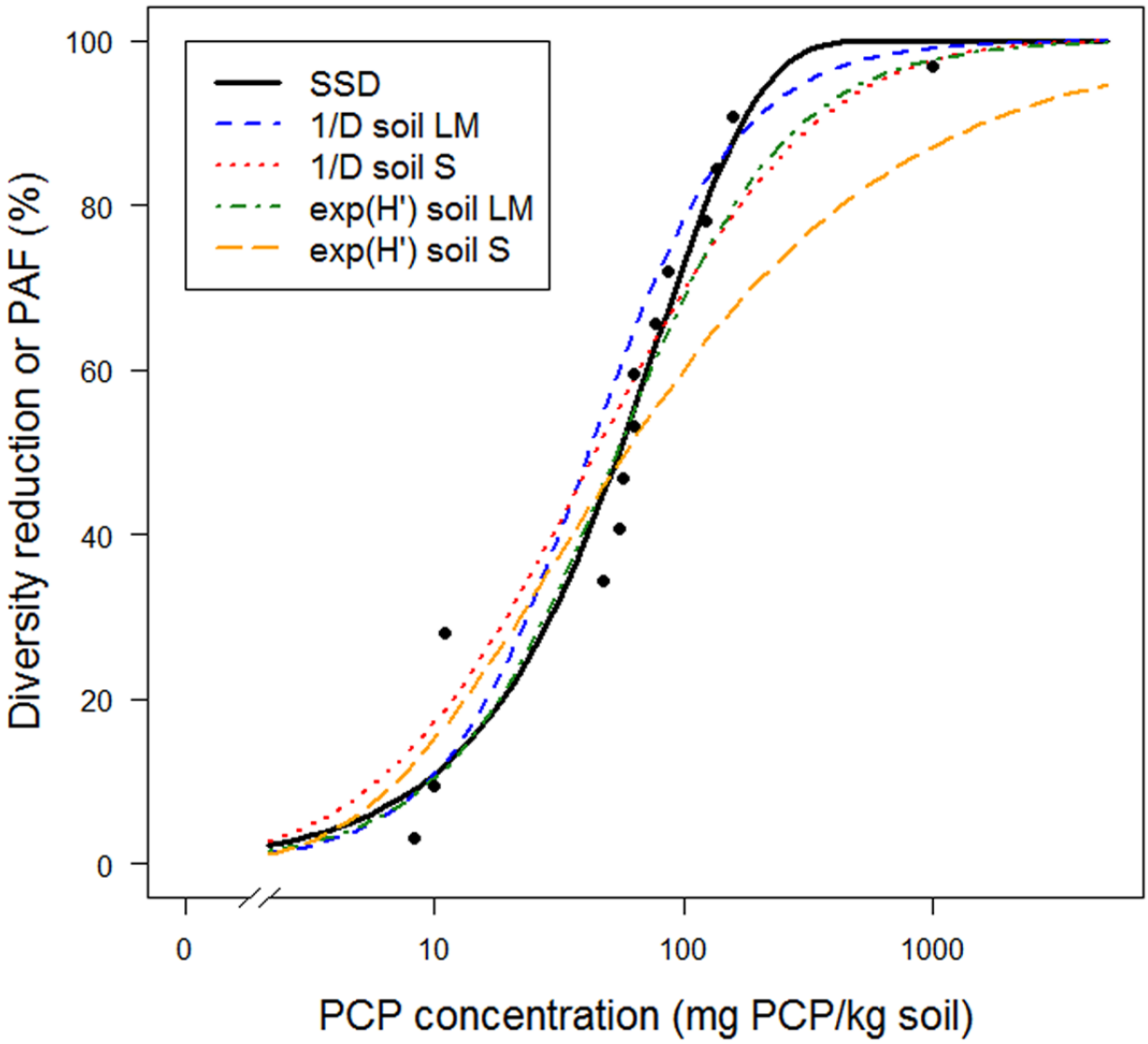

Figure 6-3 : Species sensitivity distribution produced for PCP in terrestrial ecosystems. The SSD curve was drawn from the geometric mean of the EC_{50} available for the different species belonging to four taxonomic groups (Tableau 6-4). The black dots correspond to the EC_{50} of these 14 species.

6.2.5 Discussion

6.2.5.1 Using diversiy indices to model the impact of PCP towards bacterial communities

The objective of our work was to evaluate the usefulness of diversity indices to measure the impact of PCP on soil bacterial communities. Even though it is difficult to generalize for other contaminants, the results obtained here for PCP are very promising. The three diversity indices calculated in this study were between 10 and over 40 times more sensitive to PCP than the FDA hydrolysis assay.

The diversity-based EC_{50} in our study can also be compared to those produced by other authors (Martí et al., 2011). These authors obtained EC_{50} ranging between 263 and 588 mg/kg for cumulative soil respiration and between 52 and 690 mg/kg for substrate-induced respiration. From their DGGE migration profiles, they visually determined that the community structure began to change at concentrations between 10 and 100 mg/kg. The diversity-based EC_{50} presented here are lower than these functional EC_{50} (Tableau 6-3), except for the value of 52 mg/kg. The EC_{50} of 52 mg/kg produced by Martí et al. (2011) corresponded to a soil with a pH of 8.0 and an organic carbon content of 1.7% (w/w). The aqueous solubility of PCP is strongly affected by pH (Arcand et al., 1995), and organic carbon plays an important role in PCP adsorption in soils (Cea et al., 2007). Considering that PCP bioavailability in the soil used by Martì et al. was probably particularly high, it is not surprising that the EC_{50} produced for this soil was low.

Another element to consider in order to recommend the use of diversity-based indicators for toxicity modeling is the robustness of the indices. The structure-based indices like $\exp^{H'}$ and especially 1/D are much more robust than community richness (Haegeman et al., 2013) and can be seen as suitable for C-R modeling. With its low dependence towards richness, high robustness and high sensitivity to changes in community structure, 1/D was deemed to be the most satisfactory index. A point that must be addressed is the fact that the diversity of the most contaminated samples (LM_7 and S_7) was higher than that of the less contaminated ones. One possible explanation is that the PCP affected all the bacterial species, thus hampering the apparition of bright bands on the DGGE profiles while causing a significant decrease in microbial biomass (~50% if the FDA assay results are considered). Considering the effect of pH on the aqueous solubility of PCP (Arcand et al., 1995), it is not surprising that this effect appeared at lower PCP concentrations for soil S than

for soil LM (Tableau 6-1). Another element that could potentially explain these observations is the persistence of DNA in soils, which makes dead organisms detectable by PCR-DGGE (Marschner et al., 2001). In fact, all the profiles probably contain the DNA of dead organisms. However, since PCP induces bacterial mortality, it is highly plausible that the most contaminated samples contain a higher proportion of DNA from dead organisms. This can artificially inflate the apparent diversity of the bacterial communities and thus reduce the sensitivity of the structure-based metrics. It would be interesting to verify whether treating the samples with ethidium monoazide bromide prior to DNA extraction (Pisz, Lawrence, Schafer, and Siciliano, 2007) or amplifying RNA instead of DNA (Pennanen et al., 2004)—two methods proposed to target the active members of the bacterial communities—would yield different results.

6.2.5.2 Using the diversity of soil bacterial communities to derive EQCs

Most importantly, the results produced using the SSD modeling process must be put into perspective. The data that were used corresponded to many different endpoints (growth, reproduction, survival, etc.). As presented by Roman et al. (2007), the sensitivity of a species to xenobiotics is highly dependent on the endpoint that is considered. Furthermore, considering that the required information was not often reported (soil pH and organic matter content), the ecotoxicological dataset was not normalized for PCP bioavailability. As presented by Puglisi et al. (2009), PCP adsoprtion has the potential to significantly change its ecotoxicological impact. Consequently, the standard deviations of the EC_{50} available for certain species included in Tableau 6-4 were high. Also, since the toxicity of xenobiotics varies from one taxonomic group to another (Forbes and Calow, 2002), it is important that the dataset used for SSD modeling be representative of the modeled ecosystem. In the present case, worms and plants dominated the dataset even though these groups may not naturally dominate terrestrial ecosystems. Therefore, the geometric mean of the EC_{50} for the taxonomic groups was used instead of the EC_{50} of the single species. This modeling choice moved the SSD slightly to the left of the *x-axis*, even though the correspondence between the SSD and 14 EC_{50} is still very good (Figure 6-3).

A question regarding the environmental relevance of the diversity-based indicators should also be addressed. Indeed, for the higher organisms generally used for toxicity testing, the link between the measured endpoint (mortality, reproduction, growth, etc.) and the species' ecological function is probably a lot more straightforward than in the case of bacterial communities. Even though the

link between bacterial (more generally microbial) structural diversity, soil functions and ecosystem stability (Girvan et al., 2005; Wittebolle et al., 2009) are well recognized, it is not direct (Deng, 2012; Griffiths and Philippot, 2013). It is therefore unclear how a decrease in 1/D or $exp^{H'}$ should be interpreted. However, similar questions can also be raised for the potentially affected fraction of species (PAF). Indeed, SSD modeling uses ecotoxicological data produced in controlled environments, often recorded for different endpoints, without considering the intra- and inter-species interactions that could potentially modulate the impact of contaminants on real ecosystems.

Though presented as EC_x, the ecotoxicological data produced in this study should be seen as a community response to PCP rather than single species toxicity indicators. Considering that the curves presented in Figure 6-3 were the same for all the percentiles between 5 and 50, it can be concluded that, for the samples studied in our work, the sensitivity of soil bacterial communities to PCP is not different from the sensitivity of terrestrial communities composed of higher species exposed to the same contaminant. It is also interesting to compare the diversity-based C-R curves with the soil screening levels adopted by various regulatory agencies for PCP. In the United States, threshold values of 5 mg/kg for terrestrial plants and 31 mg/kg for invertebrates were adopted (U.S. Environmental Protection Agency, 2007). In Canada, the soil quality guideline values are 11 mg/kg for environmental protection and 7.6 mg/kg for human health protection (Canadian Council of Ministers of the Environment, 1999). In the Netherlands, a similar intervention value of 12 mg/kg was calculated (Ministry of Housing Spatial Planning and the Environment, 2009). These threshold values are all very similar to the 10th percentile of the 1/D or $exp^{H'}$ curves produced in this study. These results are highly promising considering that SSD curves and EQCs are generated using dozens of ecotoxicological data on various higher species while the C-R curves presented here were produced by a single experiment that used small amounts of contaminated soils.

In conclusion, the use of structure-based metrics for toxicity testing presents indisputable strengths. The production of the diversity-based indicators presented in this paper required small amounts of contaminated soils, relied on a cost-effective methodology and focused on organisms that were not grown in a laboratory but indigenous to the tested environments. Furthermore, unlike traditional SSD modeling, the assessment of the diversity of soil bacterial communities implicitly considers the intra- and inter-species interactions. Of course, the results produced here for PCP cannot be generalized to other contaminants. If similarities in the sensitivity of bacterial and higher species communities can be established for other contaminants or classes of contaminants this

methodology has the potential to help fill many gaps in terrestrial ecotoxicological databases and the understanding of the impact of xenobiotics (or complex mixtures) on terrestrial ecosystems. If the similarities observed here for PCP are fortuitous, the framework still has the potential to provide relevant information on the toxicity of xenobiotics to organisms virtually absent from the ecotoxicological databases: the soil bacterial, archeal and fungal communities.

6.2.6 **Acknowledgements**

The authors would like to sincerely thank Lucie Jean for her precious help with the laboratory work carried out to produce this paper. The authors also acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada. The CIRAIG would also like to thank its industrial partners for their financial support: ArcelorMittal, Bombardier, Bell Canada, Cascades, Eco Entreprises Québec, RECYC-QUÉBEC, Groupe EDF, Gaz de France, Hydro-Québec, Johnson & Johnson, LVMH, Michelin, Mouvement des caisses Desjardins, Nestlé, Rio Tinto Alcan, RONA, SAQ, Solvay, Total, Umicore, and Veolia Environment.

CHAPITRE 7 DISCUSSION GÉNÉRALE

La méthode d'élongation présentée dans cette thèse s'est avéré être un outil très puissant afin d'estimer la diversité bactérienne des sols étudiés. Effectivement, bien que produits à partir d'une technique étant préalablement reconnue pour générer de l'information uniquement sur les espèces le plus abondantes des communautés étudiées, les estimés de diversité produits à partir du DGGE pour le second manuscrit ont été jugés plus fiables que ceux produits à partir de l'Ion Torrent PGM. Par contre, il est tout de même important de mettre ces résultats en perspectives.

D'abord, le PSR est un paramètre ayant une très grande influence sur les indices de diversité estimés à partir des profils DGGE. Bien que l'approche utilisée afin d'analyser les profils ait été en mesure d'estimer ce paramètre avec une précision de ± 2,5% pour tous les profils *in silico* analysés pour le premier manuscrit, il est très probable que la même précision ne puisse pas être atteinte lorsque des profils expérimentaux sont analysés. Effectivement, le passage du produit PCR dans le gel DGGE est susceptible de générer un bruit de fond qui peut difficilement être séparé du signal laissé par les espèces ciblées par les amorces PCR. Aussi, les conditions menant à la production des gels DGGE auront une influence sur les résultats produits. Afin de tracer un profil de bruit de fond aussi précis que possible, il est primordial que le gradient dénaturant choisi afin de produire le gel permette l'inclusion de toutes les bandes DGGE. La même remarque s'applique lorsque le gel est photographié; le zoom choisi doit inclure toutes les bandes. Une attention particulière doit également être portée lorsque le gel DGGE est préparé afin d'obtenir les profils les plus droits possible (absence de « smilings » et autres imperfections). Bien qu'elles puissent être partiellement corrigées, ces imperfections modifient la forme des pics DGGE et compliquent l'analyse des profils. Bien entendu, ces éléments ont tous la capacité de modifier à la hausse ou à la baisse les estimés de diversité produits par la méthode d'élongation. Par contre, lorsque cette dernière est utilisée à des fins comparatives et que les échantillons comparés sont injectés sur le même gel DGGE, il est probable que l'influence de ces éléments sera similaire pour chaque échantillon.

Aussi, les observations faites dans le cadre du premier manuscrit montrent que les patrons de dominance extraits des profils DGGE sont semblables à un groupement des séquences d'ADN à une similitude de 98%. Dans les deux premiers manuscrits, la longueur des séquences d'ADN analysées a peu varié. Effectivement, les séries de données utilisées dans le premier manuscrit

contenaient des séquences d'une longueur variant entre 200 et 300 paires de bases tandis que les amorces PCR employées dans le cadre du second manuscrit ont ciblé une région entre 130 et 170 paires de bases. Il a été montré que la longueur des séquences considérées avait un impact sur les estimés de diversité produits par les plateformes NGS (Morales, Cosart, Johnson, and Holben, 2009). Il est donc possible que l'utilisation d'amorces PCR différentes, ciblant des sections plus longues ou plus courtes du gène 16S, modifie la correspondance ayant été trouvée entre le processus de (co)migration des séquences d'ADN sur les gels DGGE et celui du groupement des séquences à un pourcentage de similitude donné. De plus, en produisant les profils DGGE *in silico*, une séquence représentative unique a été sélectionnée pour chaque OTU. Dès lors, sur ces profils de migration, une espèce (OTU) était à l'origine d'une bande DGGE. Étant donné la variation intra espèce du gène 16S bactérien, une espèce unique peut produire plus d'une bande DGGE (Dowd et al., 2008). Dépendamment de la fréquence d'occurrence de ce phénomène et du degré de similitude entre le gène 16S des bactéries faisant partie d'une même espèce, les indices de diversité produits par le DGGE et les NGS pourraient diverger. Ce phénomène peut, en partie du moins, expliquer le fait que les droites de régression produites pour le second manuscrit n'avaient pas des pentes unitaires.

L'incapacité à produire des droites de régression ayant des pentes unitaires s'explique probablement également en considérant les distorsions identifiées dans les séries de données produites à l'aide de l'Ion Torrent PGM. En observant le Tableau 5-4, il est possible de constater que la pente des droites de régression produites à l'aide de l'indice de Simpson (1/D) sont bien plus près de l'unité lorsque les indices Ion Torrent sont corrigés. Puisqu'il n'a pas été possible d'estimer quantitativement, et donc de corriger, l'influence des erreurs de séquençage sur la richesse des communautés, la pente des droites de régression correspondant à la richesse et, dans une moindre mesure, à l'indice de Shannon ($\exp^{H'}$), ne se sont jamais rapprochées de l'unité. Au contraire, le fait de corriger les indices Ion Torrent afin de tenir compte de la faible profondeur de séquençage a fait passer les pentes initialement situées entre 0,93 et 1,27 à des valeurs situées entre 0,51 et 0,81. Tel que mentionné dans le second manuscrit, l'élément le plus important afin d'être en mesure d'utiliser l'une ou l'autre des techniques afin de comparer la diversité des communautés bactériennes des sols est la corrélation entre les estimés de diversité produits par les deux approches. En ce sens, toutes les corrélations furent significatives après avoir corrigé les séries de données.

Le paragraphe précédent, ainsi que le second manuscrit, dressent un portrait peu reluisant des NGS. Il est donc très important de nuancer ces propos. Les plateformes NGS sont des outils excessivement puissants afin d'étudier la diversité ainsi que la phylogénie microbienne des sols ou de tout autre écosystème. Le nombre, la qualité et la longueur des séquences lues par ces appareils augmentent de façon constante. Les difficultés rencontrées dans le second manuscrit ont été partiellement causées par la technologie Ion Torrent, qui n'est certainement pas aussi mature que d'autres plateformes (i.e. 454 pyrosequencing) mais aussi parce que la longueur des amplicons dépassait légèrement, lorsque les amorces sont considérées, les capacités théoriques de l'appareil. Il est probable que toutes les plateformes NGS auraient produit des résultats semblables si elles avaient été utilisées à la limite de leurs capacités. Lorsque la longueur des amplicons se situe en-deçà des limites de l'appareil utilisé, l'effet du pourcentage GC ou de la longueur des séquences sur l'abondance relative des OTU est probablement bien moindre que ce qui a été qualitativement observé dans le second manuscrit.

Le second manuscrit a permis de montrer que l'utilisation du DGGE ou de l'Ion Torrent étaient en mesure de mener à des conclusions écologiques semblables lorsque la diversité des communautés bactériennes des sols était comparée. Dès lors, est-il possible de considérer que les valeurs obtenues sont représentatives de l'effet du PCP sur la diversité bactérienne des sols? Il est difficile de répondre à cette question. Premièrement, il a été montré que le protocole expérimental utilisé afin d'extraire l'ADN contenu dans le sol (Carrigg, Rice, Kavanagh, Collins, and O'Flaherty, 2007; Pan et al., 2010) ou d'amplifier cet ADN (Kurata et al., 2004; Takahiro, 2003) pouvaient avoir un effet sur les profils microbiens générés à partir des méthodes moléculaires. Aussi, les amorces PCR utilisées afin d'amplifier l'ADN extrait, bien que désignées par le qualificatif « universelles », ne ciblent pas la totalité des espèces. Ainsi, peu importe quelle méthode est utilisée afin d'analyser l'ADN amplifié, le portrait des communautés ne pourra jamais être plus représentatif que le produit PCR analysé. Il est donc primordial d'optimiser les protocoles d'extraction et d'amplification d'ADN employés au laboratoire. Par contre, même si la diversité estimée à partir des approches moléculaires n'est pas quantitativement représentative de la diversité réelle des communautés étudiées, cela ne rend pas caduque l'utilisation d'un indicateur écotoxicologique basé sur la diversité bactérienne des sols. Premièrement, les amorces PCR universelles ne ciblent pas toutes les espèces mais elles en ciblent tout de même une très grande proportion. De plus, les espèces qui seront ciblées et celles qui ne le seront pas ne sont pas choisies par l'expérimentateur sur la base

de leur sensibilité à un certain contaminant ou de leur importance fonctionnelle. Ce dernier choisit simplement des amorces offrant une bonne couverture pour le groupe taxonomique étudié. En ciblant autant d'espèces choisies aléatoirement, l'indicateur a tout de même le potentiel d'être représentatif de la communauté étudiée (Forbes and Calow, 2002).

Une question ayant des implications beaucoup plus importantes par rapport à l'utilisation d'un indicateur basé sur la diversité bactérienne des sols concerne l'origine de l'ADN analysé. Effectivement, les manuscrits 2 et 3 ont montré que la diversité des échantillons contaminés à une concentration de ~2000 mg/kg était plus élevée que celle mesurée pour des échantillons contenant ~425 mg/kg de PCP. Afin d'expliquer cette observation, deux hypothèses ont été formulées. Premièrement, la très forte contamination a pu avoir un effet toxique sur toutes les espèces bactériennes, faisant en sorte que le profil migratoire DGGE a peu changé (S_7 vs S_0). Cette hypothèse est partiellement appuyée par le fait que la biomasse microbienne a été significativement réduite par le PCP (~ 50% en considérant l'activité hydrolytique de la FDA). Une autre hypothèse permettant d'expliquer cette observation est que l'ADN des bactéries tuées par l'ajout de PCP n'a pas pu être complètement dégradé en 28 jours. En fait, les deux phénomènes se sont probablement produits simultanément. À forte concentration, le PCP a probablement induit une forte mortalité et inhibé la croissance de toutes les espèces bactériennes. Si l'inhibition de la croissance des espèces bactériennes constitue un effet réel du PCP, l'inclusion de l'ADN appartenant à des organismes morts est un artéfact des études basées sur l'extraction de l'ADN contenu dans le sol. Cet artéfact a le potentiel de gonfler les estimés de diversité produits à partir du DGGE ou des NGS. De plus, l'effet quantitatif de cet artéfact a probablement augmenté en parallèle avec la concentration en PCP ajoutée aux sols, modifiant ainsi la forme des courbes concentration-réponse tracées à partir des indices de diversité. Tel que discuté dans le troisième manuscrit, il serait intéressant de quantifier l'effet de cet artéfact en traitant les sols avec du bromure d'éthydium monoazide (Pisz et al., 2007) ou encore en amplifiant l'ARN plutôt que l'ADN (Pennanen et al., 2004). Ces deux méthodes permettent de cibler les membres actifs des communautés bactériennes.

Malgré toutes les distorsions mentionnées précédemment, inhérentes au modèle d'élongation présenté dans cette thèse mais également aux plateformes NGS, l'utilité des indices de diversité dans le domaine de l'écotoxicologie terrestre mérite plus d'attention. Bien entendu, les mesures fonctionnelles d'activité microbienne ou enzymatique présentent un intérêt certain lorsque l'on s'intéresse à l'état général d'un sol. Par contre, si l'objectif est de proposer un indicateur pouvant

potentiellement être normalisé qui permet de mesurer l'effet d'un contaminant ou de comparer l'effet de plusieurs contaminants sur la communauté bactérienne d'un sol, les indices de diversité présentent un potentiel supérieur aux indicateurs fonctionnels. Tel que mentionné précédemment, les indices de diversité permettent de fractionner les communautés bactériennes en unités taxonomiques plutôt que de considérer ces dernières comme un tout homogène. Tel qu'anticipé, le fait de fractionner les communautés bactériennes en plusieurs sous-groupes a permis aux indicateurs basés sur la diversité 1) d'être plus sensibles au PCP qu'un indicateur fonctionnel et 2) de ne pas augmenter suite à l'ajout de contaminant dans le sol. Une autre difficulté liée à l'utilisation de mesures fonctionnelles afin d'évaluer la toxicité des contaminants envers les communautés bactériennes des sols est de choisir quelle(s) fonction(s) doit/doivent être mesurée(s). En effet, dépendamment de son mode d'action toxique ou de sa conformation spatiale, un contaminant risque d'affecter certaines fonctions du sol plus que d'autres. La perte de diversité taxonomique semble être une mesure permettant de comparer l'effet de contaminants ayant des modes d'action différents. De plus, il s'agit d'une mesure qui s'intègre très bien à la modélisation SSD, qui cherche à quantifier la proportion des espèces d'un écosystème affectées par un contaminant plutôt que de prédire une perte de fonction de l'écosystème.

La similitude observée dans le troisième manuscrit entre les courbes concentration-réponse tracées à partir des indices de Shannon ($exp^{H'}$) ou de Simpson (1/D) et la SSD est un résultat présentant un fort potentiel. Il serait aisé d'affirmer que cette correspondance n'est que fortuite. Par contre, étant donné très grande diversité génétique et métabolique des communautés bactériennes présente dans les sols, il n'est pas impossible que la gamme de sensibilité des différentes espèces bactériennes face à un contaminant soit représentative de la gamme de sensibilité des espèces dites supérieures exposées au même contaminant. Dès lors, il sera intéressant de tenter d'étendre ces observations à d'autres substances que le PCP. Si une telle similitude peut être établie, pour certaines classes de contaminants ou pour les contaminants agissant selon certains modes d'action toxique, l'utilisation du modèle d'élongation présenté dans cette thèse pourrait avoir un apport majeur aux domaines de l'analyse de risque environnemental et de l'analyse du cycle de vie. Effectivement, afin de produire des résultats minimalement fiables, la modélisation SSD requiert la réalisation d'essais de toxicité sur un minimum de 10 espèces différentes. Au contraire, l'application du modèle d'élongation ne requiert que quelques grammes de sol par niveau de concentration et est basé sur une technique très répandue. Et si une similitude entre les courbes concentration-réponse basées sur les indices

de diversité et les SSD ne peut pas être établie, les données écotoxiques produites à partir de cette approche constituent tout de même une source d'information très pertinente afin de mieux comprendre l'effet des contaminants sur des groupes taxonomiques très importants dans les écosystèmes terrestres. Dans ce projet, seules les communautés bactériennes des sols ont été ciblées. Par contre, l'approche proposée peut également s'appliquer aux communautés fongiques et archéennes des sols, si différentes amorces PCR sont choisies. L'étude simultanée de ces trois groupes permettrait d'avoir une vision plus globale de l'effet des contaminants sur les communautés microbiennes des sols. Par contre, avant d'utiliser le modèle d'élongation présenté dans cette thèse afin d'étudier ces communautés, il faudra revoir le paramétrage du modèle afin d'adapter le modèle aux communautés fongiques et archéennes.

Finalement, il importe de mettre de l'avant que l'approche employée dans ce projet n'est pas le seul moyen d'accroître la disponibilité des données écotoxiques terrestres. Effectivement, une certaine quantité d'indicateurs écotoxiques concernant des espèces de mammifères et d'oiseaux sont également disponibles. Par contre, ces indicateurs ne sont pas utilisables avec les données concernant les espèces à contact direct. Il serait possible de développer une approche permettant d'utiliser simultanément les deux types de données, non seulement pour augmenter la disponibilité des données mais aussi pour accroître la représentativité des séries de données écotoxiques terrestres utilisées afin d'ajuster des courbes SSD.

CONCLUSION

Les travaux présentés dans cette thèse portaient sur l'étude de l'effet du pentachlorophénol sur la diversité des communautés bactériennes des sols à partir de la technique de l'électrophorèse sur gel en gradient dénaturant. Le projet s'appuyait sur l'hypothèse de recherche suivante : *La technique de l'électrophorèse sur gel en gradient dénaturant peut permettre la production d'un indicateur écotoxique quantitativement représentatif de l'effet du pentachlorophénol sur la diversité taxonomique des bactéries du sol.*

La méthode d'élongation des RAD DGGE a permis de produire des estimés de diversité précis lorsque des profils de migration *in silico* ont été étudiés. De plus, les estimés de diversité produits à l'aide de cette approche furent significativement corrélés à ceux produits par l'Ion Torrent. Dès lors, il est possible de conclure que les indicateurs écotoxiques présentés dans cette thèse sont aussi représentatifs que l'ont été les étapes d'extraction et d'amplification d'ADN. Cette distinction est nécessaire et s'applique à toutes les approches basées sur l'étude et l'amplification de l'ADN extrait du sol. De plus, l'utilité d'indicateurs basés sur la diversité bactérienne des sols a été testée. Ces indicateurs se sont avérés beaucoup plus sensibles au PCP que certains indicateurs fonctionnels. De plus, les courbes concentration-réponse tracées à partir des indices de diversité se sont avérées très semblables à la SSD représentative de la sensibilité des communautés biotiques terrestres exposées au PCP. Si les observations faites pour le PCP peuvent être généralisées, ces indicateurs pourraient jouer un rôle central dans la génération de connaissances quant à l'impact des contaminants (ou des mélanges complexes) sur les écosystèmes terrestres.

Les sections suivantes résument les différentes contributions apportées par les travaux présentés dans cette thèse et donnent quelques orientations afin de poursuivre ces travaux.

Contributions scientifiques

La présente thèse a permis :

- De montrer que l'approche « Rolling disk » (ou « Rolling ball ») n'est pas adaptée à l'analyse de profils DGGE complexes;

- De montrer qu'il est possible d'extraire des profils DGGE de l'information quant à la richesse et l'équité des communautés étudiées, et que cette information peut être utilisée afin d'estimer la diversité bactérienne à partir de la technique du DGGE;
- De montrer que l'injection d'une trop grande quantité d'ADN dans les profils DGGE risquait de saturer le gel localement;
- De montrer que l'Ion Torrent risquait de générer des distorsions intra-échantillon dans l'abondance relative des OTU;
- De montrer la sensibilité d'un indicateur basé sur la diversité bactérienne face au PCP;
- De faire un rapprochement entre la sensibilité des bactéries exposées au PCP et celle des communautés biotiques macroscopiques terrestres.

Contributions pratiques

Différents outils ont également été développés :

- Un cadre théorique permettant d'extraire des profils DGGE de l'information quantitativement représentative des communautés étudiées;
- Une méthode permettant d'estimer la diversité bactérienne à partir du DGGE;
- Des outils informatiques permettant l'analyse des gels DGGE à l'aide du logiciel Matlab (annexe 2);
- Des données écotoxiques représentatives de l'effet du PCP sur la diversité bactérienne des sols.

Perspectives et recommandations

La méthode d'élongation des RAD DGGE présentée dans cette thèse a permis de produire des courbes concentration-réponse représentatives de l'effet du pentachlorophénol sur la diversité des communautés bactériennes des sols. Cet indicateur s'est avéré être sensible à ce contaminant et a fait preuve d'un potentiel énorme en produisant des indicateurs quantitativement similaires aux critères de qualité environnementale en vigueur dans plusieurs pays pour cette substance. Les

recommandations qui suivent énoncent quelques perspectives de développement et de validation de cet outil.

- *Influence de la longueur des séquences amplifiées sur les indices de diversité estimés à partir du DGGE*

Dans cette thèse, il a été montré que la migration et la superposition des bandes d'ADN sur les gels DGGE étaient similaires à un groupement des séquences d'ADN à une similitude de 98%. Puisqu'il a été montré que la longueur des séquences amplifiées influençait les estimés de diversité calculés à partir des résultats des plateformes NGS, il se pourrait que ce paramètre influence également le processus de formation des profils DGGE. Ainsi, il est possible que la similitude observée entre bandes DGGE et groupement des séquences à 98% ne soit pas valide pour des séquences plus longues ou plus courtes. Le paramétrage du modèle d'élongation des RAD DGGE pourrait donc être à revoir.

- *Influence du processus d'analyse et de groupement des séquences sur les études comparatives de diversité*

La méthodologie utilisée afin d'analyser les résultats des NGS (algorithme de groupement, pourcentage de similitude, algorithmes de détection et suppression des erreurs et des chimères) a une grande influence sur les estimés de diversité produits à l'aide de ces technologies. Dans le but d'accroître l'applicabilité et l'acceptabilité de la méthode d'élongation des RAD DGGE pour les études comparatives de diversité, il serait intéressant de comprendre l'influence quantitative de la méthodologie d'analyse des données de séquençage non pas sur les estimés de diversité eux-mêmes mais bien sur la comparaison de ces derniers.

- *Détermination de la « provenance » de l'ADN analysé à partir des approches moléculaires*

Dans cette thèse, il a été observé que la diversité bactérienne des sols diminuait à mesure que la concentration en pentachlorophénol augmentait mais qu'à partir d'une certaine concentration, la diversité semblait augmenter. Pour un des deux sols utilisés, le profil DGGE correspondant à l'échantillon le plus contaminé ressemblait étrangement au profil DGGE de l'échantillon propre. Dans le but de proposer l'utilisation de tests de toxicité de routine basés sur la diversité des communautés bactériennes dans les sols, il serait très important de déterminer la cause de ces augmentations de diversité à fortes concentrations. L'ADN

provenant d'organismes morts a le potentiel de générer des distorsions dans les estimés de diversité produits, qui risquent d'augmenter avec le niveau de contamination. Dans tous les cas, il serait important de proposer une approche méthodologique permettant de cibler les microorganismes vivants.

- *Généralisation des observations faites sur le pentachlorophénol sur d'autres contaminants*

Les indicateurs écotoxicologiques basés sur les indices de diversité produits dans cette thèse se sont avérés sensibles au pentachlorophénol et quantitativement similaires aux critères de qualité environnementale en vigueur pour ce contaminant dans plusieurs pays du monde. La recommandation principale devant ressortir de ce travail de recherche est d'étudier l'effet de plus de contaminants sur la diversité bactérienne des sols. Si les observations faites pour le pentachlorophénol sont généralisables à d'autres contaminants, ou encore à certaines classes de substances, l'approche proposée dans ce projet pourrait devenir un outil excessivement puissant afin d'étudier l'effet des contaminants sur les écosystèmes terrestres.

- *Augmentation de la disponibilité des données écotoxiques terrestres*

Bien que le présent projet se soit concentré sur la génération de nouvelles données écotoxiques, il a été observé que plusieurs données actuellement disponibles ne sont pas utilisables. Il serait donc intéressant de développer des modèles conceptuels représentatifs de certains écosystèmes d'intérêts (agroécosystème, forestier, etc.) afin de faire le lien entre une concentration en contaminant dans le sol et la dose ingérée par certains mammifères et oiseaux. Ainsi, il sera possible d'intégrer les données écotoxiques représentatives des mammifères et des oiseaux aux données représentatives des organismes à contact direct.

BIBLIOGRAPHIE

Adam, G, and Duncan, H. (2001). Development of a sensitive and rapid method for the measurement of total microbial activity using fluorescein diacetate (FDA) in a range of soils. *Soil Biology and Biochemistry, 33*(7–8), 943-951. doi: http://dx.doi.org/10.1016/S0038-0717(00)00244-3

Aldenberg, T, and Jaworska, JS. (2000). Uncertainty of the Hazardous Concentration and Fraction Affected for Normal Species Sensitivity Distributions. *Ecotoxicology and Environmental Safety, 46*, 1-18.

Arcand, Y, Hawari, J, and Guiot, SR. (1995). Solubility of pentachlorophenol in aqueous solutions: The pH effect. *Water Research, 29*(1), 131-136. doi: http://dx.doi.org/10.1016/0043-1354(94)E0104-E

ASTM International. (2006). ASTM D1140-00 - Standard Test Methods for Amount of Material in Soils Finer than No. 200 (75-µm) Sieve. West Conshohocken, PA.

ASTM International. (2007a). ASTM D422-63 - Standard Test Method for Particle-Size Analysis of Soils. West Conshohocken, PA.

ASTM International. (2007b). ASTM D2974-07a - Standard Test Methods for Moisture, Ash, and Organic Matter of Peat and Other Organic Soils. West Conshohocken, PA.

ASTM International. (2007c). ASTM Standard D4972-01 - Standard Test Method for pH of Soils. West Conshohocken, PA.

ASTM International. (2012). ASTM E1676 - 12 Standard Guide for Conducting Laboratory Soil Toxicity or Bioaccumulation Tests with the Lumbricid Earthworm Eisenia Fetida and the Enchytraeid Potworm Enchytraeus albidus. West Conshohocken, PA.

Baillie, JEM, and Upham, K. (2013). Species Diversity Within and Among Ecosystems. In R. Leemans (Ed.), *Ecological Systems* (pp. 257-271): Springer New York.

Becaert, V, Samson, R, and Deschenes, L. (2006). Effect of 2,4-D contamination on soil functional stability evaluated using the relative soil stability index (RSSI). *Chemosphere, 64*(10), 1713-1721.

Blackwood, CB, Hudleston, D, Zak, DR, and Buyer, JS. (2007). Interpreting Ecological Diversity Indices Applied to Terminal Restriction Fragment Length Polymorphism Data: Insights from Simulated Microbial Communities. *Applied and Environmental Microbiology, 73*(16), 5276-5283. doi: 10.1128/aem.00514-07

Boutin, C, and Rogers, CA. (2000). Pattern of Sensitivity of Plant Species to Various Herbicides--An Analysis with Two Databases. *Ecotoxicology, 9*(4), 255.

Boyle, D. (2006). Effects of pH and cyclodextrins on pentachlorophenol degradation (mineralization) by white-rot fungi. *Journal of Environmental Management, 80*(4), 380-386. doi: 10.1016/j.jenvman.2005.09.017

Buée, M, Reich, M, Murat, C, Morin, E, Nilsson, RH, Uroz, S, and Martin, F. (2009). 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist, 184*(2), 449-456. doi: 10.1111/j.1469-8137.2009.03003.x

Bürgmann, H, Pesaro, M, Widmer, F, and Zeyer, J. (2001). A strategy for optimizing quality and quantity of DNA extracted from soil. *Journal of Microbiological Methods, 45*(1), 7-20. doi: 10.1016/s0167-7012(01)00213-5

Butchart, SHM, Walpole, M, Collen, B, van Strien, A, Scharlemann, JPW, Almond, REA, . . . Watson, R. (2010). Global Biodiversity: Indicators of Recent Declines. *Science, 328*(5982), 1164-1168. doi: 10.1126/science.1187512

Canadian Council of Ministers of the Environment. (1999). *Canadian Soil Quality Guidelines for the Protection of Environmental and Human Health; Pentachlorophenol (1997)*. Winnipeg: Retrieved from http://ceqg-rcqe.ccme.ca/download/en/276/.

Carrigg, C, Rice, O, Kavanagh, S, Collins, G, and O'Flaherty, V. (2007). DNA extraction method affects microbial community profiles from soils and sediment. *Applied Microbiology and Biotechnology, 77*(4), 955-964.

Cea, M, Seaman, JC, Jara, AA, Fuentes, B, Mora, ML, and Diez, MC. (2007). Adsorption behavior of 2,4-dichlorophenol and pentachlorophenol in an allophanic soil. *Chemosphere, 67*(7), 1354-1360. doi: http://dx.doi.org/10.1016/j.chemosphere.2006.10.080

Cleland, EE. (2012). Biodiversity and Ecosystem Stability. *Nature Education Knowledge, 3*(10), 14.

Cline, J, Braman, JC, and Hogrefe, HH. (1996). PCR Fidelity of Pfu DNA Polymerase and Other Thermostable DNA Polymerases. *Nucleic Acids Research, 24*(18), 3546-3551. doi: 10.1093/nar/24.18.3546

Cole, JR, Chai, B, Farris, RJ, Wang, Q, Kulam, SA, McGarrell, DM, . . . Tiedje, JM. (2005). The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Research, 33*(suppl 1), D294-D296. doi: 10.1093/nar/gki038

Cole, JR, Wang, Q, Cardenas, E, Fish, J, Chai, B, Farris, RJ, . . . Tiedje, JM. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research, 37*(1), 141-145. doi: 10.1093/nar/gkn879

Demanou, J, Sharma, S, Dorfler, U, Schroll, R, Pritsch, K, Njine, T, . . . Schloter, M. (2006). Structural and functional diversity of soil microbial communities as a result of combined applications of copper and mefenoxam. *Soil Biology and Biochemistry, 38*(8), 2381-2389.

Demidenko, E, Williams, BB, Flood, AB, and Swartz, HM. (2012). Standard error of inverse prediction for dose–response relationship: approximate and exact statistical inference. *Statistics in Medicine*. Retrieved from http://dx.doi.org/10.1002/sim.5668 doi:10.1002/sim.5668

Deng, H. (2012). A review of diversity-stability relationship of soil microbial community: What do we not know? *Journal of Environmental Sciences, 24*(6), 1027-1035. doi: http://dx.doi.org/10.1016/S1001-0742(11)60846-2

Díez, B, Pedrós-Alió, C, Marsh, TL, and Massana, R. (2001). Application of Denaturing Gradient Gel Electrophoresis (DGGE) To Study the Diversity of Marine Picoeukaryotic Assemblages and Comparison of DGGE with Other Molecular Techniques. *Applied and Environmental Microbiology, 67*(7), 2942-2951. doi: 10.1128/aem.67.7.2942-2951.2001

Dohm, JC, Lottaz, C, Borodina, T, and Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research, 36*(16), e105. doi: 10.1093/nar/gkn425

Doroghazi, JR, and Buckley, DH. (2008). Evidence from GC-TRFLP that Bacterial Communities in Soil Are Lognormally Distributed. *PLoS ONE, 3*(8), e2910. doi: 10.1371/journal.pone.0002910

Dowd, S, Sun, Y, Secor, P, Rhoads, D, Wolcott, B, James, G, and Wolcott, R. (2008). Survey of bacterial diversity in chronic wounds using Pyrosequencing, DGGE, and full ribosome shotgun sequencing. *BMC Microbiology, 8*(1), 43.

Dowse, R, Tang, D, Palmer, CG, and Kefford, BJ. (2013). Risk assessment using the species sensitivity distribution method: Data quality versus data quantity. *Environmental Toxicology and Chemistry, 32*(6), 1360-1369. doi: 10.1002/etc.2190

Dunbar, J, Barns, SM, Ticknor, LO, and Kuske, CR. (2002). Empirical and Theoretical Bacterial Diversity in Four Arizona Soils. *Applied and Environmental Microbiology, 68*(6), 3035-3045. doi: 10.1128/aem.68.6.3035-3045.2002

Edgar, RC, Haas, BJ, Clemente, JC, Quince, C, and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics, 27*(16), 2194-2200. doi: 10.1093/bioinformatics/btr381

Ereshefsky, M. (2010). Microbiology and the species problem. *Biology and Philosophy, 25*(4), 553-568. doi: 10.1007/s10539-010-9211-9

Feinstein, LM, Sul, WJ, and Blackwood, CB. (2009). Assessment of Bias Associated with Incomplete Extraction of Microbial DNA from Soil. *Applied and Environmental Microbiology, 75*(16), 5428-5433. doi: 10.1128/aem.00120-09

Fierer, N, Breitbart, M, Nulton, J, Salamon, P, Lozupone, C, Jones, R, . . . Jackson, RB. (2007). Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied and Environmental Microbiology, 73*(21), 7059-7066.

Filip, Z. (2002). International approach to assessing soil quality by ecologically-related biological parameters. *Agriculture, Ecosystems & Environment, 88*(2), 169-174. doi: http://dx.doi.org/10.1016/S0167-8809(01)00254-7

Forbes, VE, and Calow, P. (2002). Species sensitivity distributions revisited: A critical appraisal. *Human and Ecological Risk Assessment, 8*(3), 473-492.

Frampton, GK, Jänsch, S, Scott-Fordsmand, JJ, Römbke, J, and van den Brink, PJ. (2006). Effects of pesticides on soil invertebrates in laboratory studies: A review and analysis using species sensitivity distributions. *Environmental Toxicology and Chemistry, 25*(9), 2480.

Fuglem, M, Parr, G, and Jordaan, IJ. (2013). Plotting positions for fitting distributions and extreme value analysis. *Canadian Journal of Civil Engineering, 40*(2), 130-139. doi: 10.1139/cjce-2012-0427

Gans, J, Wolinsky, M, and Dunbar, J. (2005). Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil. *Science, 309*(5739), 1387-1390. doi: 10.1126/science.1112665

Gaudet, CL, Bright, D, Adare, K, and Potter, K. (2001). A rank-based approach to deriving Canadian soil and sediment quality guidelines. In L. Posthuma, G. W. Suter II & T. P. Traas (Eds.), *Species sensitivity distributions in ecotoxicology* (pp. 255-273). Boca Raton: Lewis Publisher.

Giddings, JM, Solomon, KR, and Maund, SJ. (2001). Probabilistic risk assessment of cotton pyrethroids: II. Aquatic mesocosm and field studies. *Environmental Toxicology and Chemistry, 20*(3), 660-668.

Girvan, MS, Campbell, CD, Killham, K, Prosser, JI, and Glover, LA. (2005). Bacterial diversity promotes community stability and functional resilience after perturbation. *Environmental Microbiology, 7*(3), 301-313.

Glenn, TC. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources, 11*(5), 759-769. doi: 10.1111/j.1755-0998.2011.03024.x

Green, V, Stott, D, and Diack, M. (2006). Assay for fluorescein diacetate hydrolytic activity: Optimization for soil samples. *Soil Biology and Biochemistry, 38*(4), 693-701. doi: doi: 10.1016/j.soilbio.2005.06.020

Griffiths, BS, and Philippot, L. (2013). Insights into the resistance and resilience of the soil microbial community. *FEMS Microbiology Reviews, 37*(2), 112-129. doi: 10.1111/j.1574-6976.2012.00343.x

Groombridge, B, and Jenkins, M. (2002). *World Atlas of Biodiversity: Earth's Living Resources in 21st Century*: University of California Press.

Haegeman, B, Hamelin, J, Moriarty, J, Neal, P, Dushoff, J, and Weitz, JS. (2013). Robust estimation of microbial diversity in theory and in practice. *The ISME Journal*. Retrieved from doi:10.1038/ismej.2013.10

Hammer, Ø, Ryan, P, and Harper, D. (2001). PAST: Paleontological Statistics software package for education and data analysis. *Palaeontologia Electronica, 4*(1), 9 pp.

Hill, TCJ, Walsh, KA, Harris, JA, and Moffett, BF. (2003). Using ecological diversity measures with bacterial communities. *FEMS Microbiology Ecology, 43*, 1-11.

Hooper, DU, Chapin, FS, Ewel, JJ, Hector, A, Inchausti, P, Lavorel, S, . . . Wardle, DA. (2005). Effects of Biodiversity on Ecosystem Functioning: A Consessus of Current Knowledge. *Ecological Monographs, 75*(1), 3-35. doi: 10.1890/04-0922

Hose, GC, and van den Brink, PJ. (2004). Confirming the species-sensitivity distribution concept for endosulfan using laboratory, mesocosm, and field data. *Archives of Environmental Contamination and Toxicology, 47*, 511-520.

Hugenholtz, P. (2002). Exploring prokaryotic diversity in the genomic era. *Genome Biology, 3*(2), 1-8. doi: doi:10.1186/gb-2002-3-2-reviews0003

Huse, SM, Welch, DM, Morrison, HG, and Sogin, ML. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology, 12*(7), 1889-1898. doi: 10.1111/j.1462-2920.2010.02193.x

Hutton, JR. (1977). Renaturation Kinetics and thermal stability of DNA in aqueous solutions of formamide and urea. *Nucleic Acids Research, 4*(10), 3537-3555. doi: 10.1093/nar/4.10.3537

Imfeld, G, and Vuilleumier, S. (2012). Measuring the effects of pesticides on bacterial communities in soil: A critical review. *European Journal of Soil Biology, 49*(0), 22-30. doi: http://dx.doi.org/10.1016/j.ejsobi.2011.11.010

İnceoğlu, Ö, Al-Soud, WA, Salles, JF, Semenov, AV, and van Elsas, JD. (2011). Comparative Analysis of Bacterial Communities in a Potato Field as Determined by Pyrosequencing. *PLoS ONE, 6*(8), e23321. doi: 10.1371/journal.pone.0023321

Jaenicke, S, Ander, C, Bekel, T, Bisdorf, R, Dröge, M, Gartemann, K-H, . . . Goesmann, A. (2011). Comparative and Joint Analysis of Two Metagenomic Datasets from a Biogas Fermenter Obtained by 454-Pyrosequencing. *PLoS ONE, 6*(1), e14519. doi: 10.1371/journal.pone.0014519

Jensen, J, Smith, SR, Krogh, PH, Versteeg, DJ, and Temara, A. (2007). European risk assessment of LAS in agricultural soil revisited: Species sensitivity distribution and risk estimates. *Chemosphere, 69*, 880-892.

Jost, L. (2006). Entropy and diversity. *Oikos, 113*(2), 363-375. doi: 10.1111/j.2006.0030-1299.14714.x

Khandelwal, G, and Bhyravabhotla, J. (2010). A Phenomenological Model for Predicting Melting Temperatures of DNA Sequences. *PLoS ONE, 5*(8), e12433. doi: 10.1371/journal.pone.0012433

Killham, K, and Staddon, WJ. (2002). Bioindicators and Sensors of Soil Health and the Application of Geostatistics. In R. G. Burns & R. P. Dick (Eds.), *Enzymes in the Environment; Activity, Ecology, and Applications* (pp. 391-405). New York: Marcel Dekker, Inc.

Kirk, JL, Beaudette, LA, Hart, M, Moutoglis, P, Klironomos, JN, Lee, H, and Trevors, JT. (2004). Methods of studying soil microbial diversity. *Journal of Microbiological Methods, 58*(2), 169-188. doi: http://dx.doi.org/10.1016/j.mimet.2004.04.006

Klindworth, A, Pruesse, E, Schweer, T, Peplies, J, Quast, C, Horn, M, and Glöckner, FO. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research, 41*(1), e1. doi: 10.1093/nar/gks808

Kooijman, SALM. (1987). A safety factor for LC50 values allowing for differences in sensitivity among species. *Water Research, 21*, 269-276.

Korbie, DJ, and Mattick, JS. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols, 3*(9), 1452-1456. doi: 10.1038/nprot.2008.133

Kunin, V, Engelbrektson, A, Ochman, H, and Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology, 12*(1), 118-123. doi: 10.1111/j.1462-2920.2009.02051.x

Kurata, S, Kanagawa, T, Magariyama, Y, Takatsu, K, Yamada, K, Yokomaku, T, and Kamagata, Y. (2004). Reevaluation and reduction of a PCR bias caused by reannealing of templates. *Applied and Environmental Microbiology, 70*, 7545 - 7549.

Lalande, J, Villemur, R, and Deschenes, L. (2013). A New Framework to Accurately Quantify Soil Bacterial Community Diversity from DGGE. *Microbial Ecology*. Retrieved from doi:10.1007/s00248-013-0230-3

Lalande, J, Yergeau, E, Greer, CW, Villemur, R, and Deschenes, L. (2013). Can DGGE and Ion Torrent sequencing yield similar quantitative conclusions when comparing the diversity of soil bacterial communities? *FEMS Microbiology Ecology (submitted)*.

Larsen, HF, and Hauschild, M. (2007). GM-Troph; A Low Data Demand Ecotoxicity Effect Indicator for Use in LCIA. *International Journal of LCA, 12*(2), 79-91.

Lazzaro, A, Schulin, R, Widmer, F, and Frey, B. (2006). Changes in lead availability affect bacterial community structure but not basal respiration in a microcosm study with forest soils. *Science of the Total Environment, 371*(1-3), 110-124.

Legendre, P, and Legendre, L. (1998). *Numerical Ecology. Second English Edition*: Elsevier.

Leštan, D, Leštan, M, Chapelle, JA, and Lamar, RT. (1996). Biological potential of fungal inocula for bioaugmentation of contaminated soils. *Journal of Industrial Microbiology, 16*(5), 286-294. doi: 10.1007/bf01570036

Leung, KMY, Grist, EPM, Morley, NJ, Morritt, D, and Crane, M. (2007). Chronic toxicity of tributyltin to develpment and reproduction of the European freshwater snail Lymnaea stagnalis (L.). *Chemosphere, 66*, 1358-1366.

Li, X, Zhang, H, Wu, M, Zhang, Y, and Zhang, C. (2008). Effect of methamidophos on soil fungi community in microcosms by plate count, DGGE and clone library analysis. *Journal of Environmental Sciences, 20*(5), 619-625.

Li, Z, Xu, J, Tang, C, Wu, J, Muhammad, A, and Wang, H. (2006). Application of 16S rDNA-PCR amplification and DGGE fingerprinting for detection of shift in microbial community diversity in Cu-, Zn-, and Cd-contaminated paddy soils. *Chemosphere, 62*(8), 1374-1380.

Liao, M, Chen, C-L, Zeng, L-S, and Huang, C-Y. (2007). Influence of lead acetate on soil microbial biomass and community structure in two different soils with the growth of Chinese cabbage (Brassica chinensis). *Chemosphere, 66*(7), 1197-1205. doi: 10.1016/j.chemosphere.2006.07.046

Life Technologies. (2012, June 5, 2012). Relentless Improvements in Throughput, Read-length and Data Quality Fuel Rapid Market Adoption of Ion PGM™ Benchtop Sequencer, *Acquire Media*. Retrieved from http://www.lifetechnologies.com/ca/en/home/about-us/news-gallery/press-releases/2012/reletless-improvemets-i-throughput.html

Life Technologies. (2013). DNA Oligo FAQ Retrieved April 5, 2013, from http://www.invitrogen.com/site/us/en/home/Products-and-Services/Product-Types/Primers-Oligos-Nucleotides/invitrogen-custom-dna-oligos/Technical-Resources-for-Oligonucleotides/DNA-Oligo-FAQ.html#5

Lock, K, De Schamphelaere, KAC, and Janssen, CR. (2002). The effect of lindane on terrestrial invertebrates. *Archives of Environmental Contamination and Toxicology, 42*, 217-221.

Loisel, P, Hamelin, J, Godon, J-J, Haegeman, B, and Harmand, J. (2009). E. Patent.

Loisel, P, Harmand, J, Zemb, O, Latrille, E, Lobry, C, Delgenès, J-P, and Godon, J-J. (2006). Denaturing gradient electrophoresis (DGE) and single-strand conformation polymorphism (SSCP) molecular fingerprintings revisited by simulation and used as a tool to measure microbial diversity. *Environmental Microbiology, 8*(4), 720-731. doi: 10.1111/j.1462-2920.2005.00950.x

Loman, NJ, Misra, RV, Dallman, TJ, Constantinidou, C, Gharbia, SE, Wain, J, and Pallen, MJ. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology, 30*(5), 434-439. doi: 10.1038/nbt.2198

Lopes, AR, Faria, C, Prieto-Fernandez, A, Trasar-Cepeda, C, Manaia, CM, and Nunes, OC. (2011). Comparative study of the microbial diversity of bulk paddy soil of two rice fields subjected to organic and conventional farming. *Soil Biology and Biochemistry, 43*(1), 115-125. doi: 10.1016/j.soilbio.2010.09.021

Lorenz, N, Hintemann, T, Kramarewa, T, Katayama, A, Yasuta, T, Marschner, P, and Kandeler, E. (2006). Response of microbial activity and microbial community composition in soils to long-term arsenic and cadmium exposure. *Soil Biology and Biochemistry, 38*(6), 1430-1437.

Loy, A, Maixner, F, Wagner, M, and Horn, M. (2007). probeBase--an online resource for rRNA-targeted oligonucleotide probes: new features 2007. *Nucleic Acids Research, 35*(Database issue), D800-804. doi: 10.1093/nar/gkl856

Macur, RE, Wheeler, JT, Burr, MD, and Inskeep, WP. (2007). Impacts of 2,4-D application on soil microbial community structure and on populations associated with 2,4-D degradation. *Microbiological Research, 162*(1), 37-45.

Magurran, AE. (2004). *Measuring biological diversity*. Oxford: Blackwell Science Ltd.

Maltby, L, Blake, N, Brock, TCM, and van den Brink, PJ. (2005). Insecticide species sensitivity distributions: Importance of test species selection and relevance to aquatic ecosystems. *Environmental Toxicology and Chemistry, 24*(2), 379.

Marschner, P, Yang, CH, Lieberei, R, and Crowley, DE. (2001). Soil and plant specific effects on bacterial community composition in the rhizosphere. *Soil Biology and Biochemistry, 33*(11), 1437-1445. doi: http://dx.doi.org/10.1016/S0038-0717(01)00052-9

Martí, E, Sierra, J, Cáliz, J, Montserrat, G, Vila, X, Garau, MA, and Cruañas, R. (2011). Ecotoxicity of chlorophenolic compounds depending on soil characteristics. *Science of the Total Environment, 409*(14), 2707-2716. doi: http://dx.doi.org/10.1016/j.scitotenv.2011.03.005

McCann, KS. (2000). The diversity-stability debate. [10.1038/35012234]. *Nature, 405*(6783), 228-233.

Megharaj, M, Singleton, I, and McClure, NC. (1998). Effect of pentachlorophenol pollution towards microalgae and microbial activities in soil from a former timber processing facility. *Bulletin of Environmental Contamination and Toxicology, 61*(1), 108-115. doi: 10.1007/s001289900736

Millenium Ecosystem Assessment. (2005). Ecosystems and Human Well-Being; Biodiversity Synthesis. Washington, DC.

Ministry of Housing Spatial Planning and the Environment. (2009). *Soil Remediation Circular 2009*. The Netherlands: Retrieved from http://www.esdat.com.au/Environmental%20Standards/Dutch/ENGELSE%20versie%20circulaire%20Bodemsanering%202009.pdf.

Morales, SE, Cosart, TF, Johnson, JV, and Holben, WE. (2009). Extensive phylogenetic analysis of a soil bacterial community illustrates extreme taxon evenness and the effects of amplicon length, degree of coverage, and DNA fractionation on classification and ecological parameters. *Applied and Environmental Microbiology, 75*(3), 668-675.

Muyzer, G, de Waal, EC, and Uitterlinden, AG. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology, 59*(3), 695-700.

Myers, RM, Maniatis, T, and Lerman, LS. (1987). Detection and localization of single base changes by denaturing gradient gel electrophoresis. In W. Ray (Ed.), *Methods in Enzymology* (Vol. Volume 155, pp. 501-527): Academic Press.

Nacke, H, Thürmer, A, Wollherr, A, Will, C, Hodac, L, Herold, N, . . . Daniel, R. (2011). Pyrosequencing-Based Assessment of Bacterial Community Structure Along Different

Management Types in German Forest and Grassland Soils. *PLoS ONE, 6*(2), e17000. doi: 10.1371/journal.pone.0017000

Nakatsu, CH. (2007). Soil microbial community analysis using denaturing gradient gel electrophoresis. *Soil Science Society of America Journal, 71*, 562-571. doi: 10.2136/sssaj2006.0080

Nannipieri, P, Ascher, J, Ceccherini, MT, Landi, L, Pietramellara, G, and Renella, G. (2003). Microbial diversity and soil functions. *European Journal of Soil Science, 54*(4), 655-670. doi: 10.1046/j.1351-0754.2003.0556.x

Narang, R, and Dunbar, J. (2004). Modeling Bacterial Species Abundance from Small Community Surveys. *Microbial Ecology, 47*(4), 396-406. doi: 10.1007/s00248-003-1026-7

NCBI. (2012). The NCBI Sequence Read Archive (SRA). Retrieved August 1st 2012

Newman, MC, Ownby, DR, Mézin, LCA, Powell, MA, Christensen, TRL, Lerberg, SB, . . . Padma, TV. (2001). Species sensitivity distributions in ecological risk assessment: Distributional assumptions, alternate bootstrap techniques, and estimation of adequate number of species. In L. Posthuma, G. W. Suter II & T. P. Traas (Eds.), *Species sensitivity distributions in ecotoxicology* (pp. 119-132). Boca Raton: Lewis Publisher.

Odum, EP. (1985). Trends Expected in Stressed Ecosystems. *BioScience, 35*(7), 419-422.

OECD. (1992). OECD Guidelines for Testing of Chemicals; Fish, Acute Toxicity Test (pp. 9): OCDE.

Office québécois de la langue française. (2012). Le grand dictionnaire terminologique Retrieved 26 mars 2013, from http://gdt.oqlf.gouv.qc.ca/index.aspx

Oros-Sichler, M, Costa, R, Heuer, H, and Smalla, K. (2007). Molecular fingerprinting techniques to analyze soil microbial communities. In J. D. van Elsas, J. K. Jansson & J. T. Trevors (Eds.), *Modern Soil Microbiology (second edition)* (pp. 355-386). Boca Raton: CRC Press.

Pan, Y, Bodrossy, L, Frenzel, P, Hestnes, A-G, Krause, S, Luke, C, . . . Bodelier, PLE. (2010). Impacts of inter- and intralaboratory variations on the reproducibility of microbial community analyses. *Applied and Environmental Microbiology, 76*(22), 7451-7458. doi: 10.1128/aem.01595-10

Paul, EA. (2007). *Soil microbiology, ecology and biochemistry (third edition)*: Elsevier Academic Press.

Pennanen, T, Caul, S, Daniell, TJ, Griffiths, BS, Ritz, K, and Wheatley, RE. (2004). Community-level responses of metabolically-active soil microorganisms to the quantity and quality of substrate inputs. *Soil Biology and Biochemistry, 36*(5), 841-848.

Pinard, R, de Winter, A, Sarkis, G, Gerstein, M, Tartaro, K, Plant, R, . . . Leamon, J. (2006). Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics, 7*(1), 216.

Pinto, AJ, and Raskin, L. (2012). PCR Biases Distort Bacterial and Archaeal Community Structure in Pyrosequencing Datasets. *PLoS ONE, 7*(8), e43093. doi: 10.1371/journal.pone.0043093

Pires, ACC, Cleary, DFR, Almeida, A, Cunha, Â, Dealtry, S, Mendonça-Hagler, LCS, . . . Gomes, NCM. (2012). Denaturing Gradient Gel Electrophoresis and Barcoded Pyrosequencing Reveal Unprecedented Archaeal Diversity in Mangrove Sediment and Rhizosphere Samples. *Applied and Environmental Microbiology, 78*(16), 5520-5528. doi: 10.1128/aem.00386-12

Pisz, JM, Lawrence, JR, Schafer, AN, and Siciliano, SD. (2007). Differentiation of genes extracted from non-viable versus viable micro-organisms in environmental samples using ethidium monoazide bromide. *Journal of Microbiological Methods, 71*(3), 312-318. doi: 10.1016/j.mimet.2007.09.015

Polz, MF, and Cavanaugh, CM. (1998). Bias in Template-to-Product Ratios in Multitemplate PCR. *Applied and Environmental Microbiology, 64*(10), 3724-3730.

Posthuma, L, Suter II, GW, and Traas, TP. (2001). *Species Sensitivity Distribution in Ecotoxicology*. Boca Raton: Lewis Publisher.

Posthuma, L, Traas, TP, and Suter II, GW. (2001). General introduction to species sensitivity distributions. In L. Posthuma, G. W. Suter II & T. P. Traas (Eds.), *Species sensitivity distributions in ecotoxicology* (pp. 3-10). Boca Raton: Lewis Publisher.

Pu, X, and Cutright, TJ. (2006). Sorption-desorption behavior of PCP on soil organic matter and clay minerals. *Chemosphere, 64*(6), 972-983. doi: 10.1016/j.chemosphere.2006.01.017

Puglisi, E, Vernile, P, Bari, G, Spagnuolo, M, Trevisan, M, Lillo, E, and Ruggiero, P. (2009). Bioaccessibility, bioavailability and ecotoxicity of pentachlorophenol in compost amended soils. *Chemosphere, 77*(1), 80-86. doi: 10.1016/j.chemosphere.2009.05.022

Qiu, X, Wu, L, Huang, H, McDonel, PE, Palumbo, AV, Tiedje, JM, and Zhou, J. (2001). Evaluation of PCR-Generated Chimeras, Mutations, and Heteroduplexes with 16S rRNA Gene-Based Cloning. *Applied and Environmental Microbiology, 67*(2), 880-887. doi: 10.1128/aem.67.2.880-887.2001

Quail, M, Smith, M, Coupland, P, Otto, T, Harris, S, Connor, T, . . . Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics, 13*(1), 341.

Quince, C, Lanzen, A, Davenport, R, and Turnbaugh, P. (2011). Removing Noise From Pyrosequenced Amplicons. *BMC Bioinformatics, 12*(1), 38.

R Developement Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rahman, MH, Okubo, A, Sugiyama, S, and Mayland, HF. (2008). Physical, chemical and microbiological properties of an Andisol as related to land use and tillage practice. *Soil and Tillage Research, 101*(1-2), 10-19.

Rajendhran, J, and Gunasekaran, P. (2011). Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiological Research, 166*(2), 99-110. doi: 10.1016/j.micres.2010.02.003

Renella, G, Mench, M, Gelsomino, A, Landi, L, and Nannipieri, P. (2005). Functional activity and microbial community structure in soils amended with bimetallic sludges. *Soil Biology and Biochemistry, 37*(8), 1498-1506.

Ritz, C. (2010). Toward a unified approach to dose–response modeling in ecotoxicology. *Environmental Toxicology and Chemistry, 29*(1), 220-229. doi: 10.1002/etc.7

Ritz, C, and Streibig, JC. (2005). Bioassay Analysis using R. *Journal of Statistical Software, 12*(5). doi: citeulike-article-id:813991

Roesch, LFW, Fulthorpe, RR, Riva, A, Casella, G, Hadwin, AKM, Kent, AD, . . . Triplett, EW. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME Journal, 1*, 283-290.

Roessink, I, Belgers, JDM, Crum, SJH, van den Brink, PJ, and Brock, TCM. (2006). Impact of triphenyltin acetate in microcosms simulating floodplain lakes. II. Comparison of species sensitivity distributions between laboratory and semi-field. *Ecotoxicology, 15*(5), 411.

Roman, YE, De Schamphelaere, KAC, Nguyen, LTH, and Janssen, CR. (2007). Chronic toxicity of copper to five benthic invertebrates in laboratory-formulated sediment: Sensitivity comparison and preliminary risk assessment. *Science of the Total Environment, 387*(1–3), 128-140. doi: http://dx.doi.org/10.1016/j.scitotenv.2007.06.023

Rosselló-Mora, R, and Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiology Reviews, 25*(1), 39-67. doi: 10.1111/j.1574-6976.2001.tb00571.x

Sambrook, J, Fritsch, EF, and Maniatis, T. (1989). *Molecular cloning: a laboratory manual*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Sanexen Services Environnementaux Inc. (2002). TerraSys (Version 1.0): Sanexen Services Environnementaux Inc.

Scelza, R, Rao, M-A, and Gianfreda, L. (2008). Response of an agricultural soil to pentachlorophenol (PCP) contamination and the addition of compost or dissolved organic matter. *Soil Biology and Biochemistry, 40*(9), 2162-2169.

Schloss, PD, Westcott, SL, Ryabin, T, Hall, JR, Hartmann, M, Hollister, EB, . . . Weber, CF. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology, 75*(23), 7537-7541. doi: 10.1128/aem.01541-09

Schmalenberger, A, and Tebbe, CC. (2003). Bacterial diversity in maize rhizospheres: conclusions on the use of genetic profiles based on PCR-amplified partial small subunit rRNA genes in ecological studies. *Molecular Ecology, 12*(1), 251-262. doi: 10.1046/j.1365-294X.2003.01716.x

Scott-Fordsmand, JJ, and Jensen, J. (2001). Ecotoxicological soil quality criteria in Denmark. In L. Posthuma, G. W. Suter II & T. P. Traas (Eds.), *Species sensitivity distributions in ecotoxicology* (pp. 275-282). Boca Raton: Lewis Publisher.

Semenzin, E, Temminghoff, EJM, and Marcomini, A. (2007). Improving ecological risk assessment by including bioavailability into species sensitivity distributions: An example for plants to nickel in soils. *Environmental Pollution, 148*, 642-647.

Sheffield, VC, Cox, DR, Lerman, LS, and Myers, RM. (1989). Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proceedings of the National Academy of Sciences, 86*(1), 232-236.

Shokralla, S, Spall, JL, Gibson, JF, and Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology, 21*(8), 1794-1805. doi: 10.1111/j.1365-294X.2012.05538.x

Sijm, DTHM, van Wezel, AP, and Crommentuijn, T. (2001). Environmental risk limits in the Netherlands. In L. Posthuma, G. W. Suter II & T. P. Traas (Eds.), *Species sensitivity distributions in ecotoxicology* (pp. 221-253). Boca Raton: Lewis Publisher.

Simon, C, and Daniel, R. (2011). Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology, 77*(4), 1153-1161. doi: 10.1128/aem.02345-10

Solomon, KR, Giddings, JM, and Maund, SJ. (2001). Probabilistic risk assessment of cotton pyrethroids: I. Distributional analyses of laboratory aquatic toxicity data. *Environmental Toxicology and Chemistry, 20*(3), 652-659.

Solomon, KR, and Takacs, P. (2001). Probabilistic risk assessment using species sensitivity distributions. In L. Posthuma, G. W. Suter II & T. P. Traas (Eds.), *Species sensitivity distributions in ecotoxicology* (pp. 285-313). Boca Raton: Lewis Publisher.

Stackebrandt, E, and Goebel, BM. (1994). Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic Bacteriology, 44*(4), 846-849. doi: 10.1099/00207713-44-4-846

Staley, JT, and Konopka, A. (1985). Measurement of in Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats. *Annual Review of Microbiology, 39*(1), 321-346. doi: doi:10.1146/annurev.mi.39.100185.001541

Stephan, CE. (2001). Use of species sensitivity distributions in the derivation of water quality criteria for aquatic life by the U.S. Environmental Protection Agency. In L. Posthuma, G. W. Suter II & T. P. Traas (Eds.), *Species sensitivity distributions in ecotoxicology* (pp. 211-220). Boca Raton: Lewis Publisher.

Stres, B, and Tiedje, J. (2006). New Frontiers in Soil Microbiology: How To Link Structure and Function of Microbial Communities? In P. Nannipieri & K. Smalla (Eds.), *Nucleic Acids and Proteins in Soil* (Vol. 8, pp. 1-22): Springer Berlin Heidelberg.

Sun, B-y, Pan, X-r, and Zhou, F. (2012). Species Sensitivity Distribution for Arsenic Toxicity on Plant Based on Soil Culture Data: Implications for Benchmarks of Soil Risk Assessments. In E. Zhu & S. Sambath (Eds.), *Information Technology and Agricultural Engineering* (Vol. 134, pp. 871-879): Springer Berlin Heidelberg.

Suter II, GW. (2001). North American history of species sensitivity distributions. In L. Posthuma, G. W. Suter II & T. P. Traas (Eds.), *Species sensitivity distributions in ecotoxicology* (pp. 11-17). Boca Raton: Lewis Publisher.

Suter II, GW, Traas, TP, and Posthuma, L. (2001). Issues and practices in the derivation and use of species sensitivity distributions. In L. Posthuma, G. W. Suter II & T. P. Traas (Eds.), *Species sensitivity distributions in ecotoxicology* (pp. 437-474). Boca Raton: Lewis Publisher.

Suzuki, M, and Giovannoni, S. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology, 62*, 625 - 630.

Swisher, R, and Carroll, G. (1980). Fluorescein diacetate hydrolysis as an estimator of microbial biomass on coniferous needle surfaces. *Microbial Ecology, 6*(3), 217-226. doi: 10.1007/bf02010387

Takahiro, K. (2003). Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering, 96*(4), 317-323. doi: 10.1016/s1389-1723(03)90130-7

Tamames, J, and Rosselló-Móra, R. (2012). On the fitness of microbial taxonomy. *Trends in Microbiology, 20*(11), 514-516. doi: http://dx.doi.org/10.1016/j.tim.2012.08.012

Temmerman, R, Huys, G, and Swings, J. (2004). Identification of lactic acid bacteria: culture-dependent and culture-independent methods. *Trends in Food Science & Technology, 15*(7–8), 348-359. doi: http://dx.doi.org/10.1016/j.tifs.2003.12.007

Thakuria, D, Schmidt, O, Mac Siurtain, M, Egan, D, and Doohan, FM. (2008). Importance of DNA quality in comparative soil microbial community structure analyses. *Soil Biology and Biochemistry, 40*, 1390-1403.

Torsvik, V, and Ovreas, L. (2007). Microbial Phylogeny and diversity in soil. In J. D. van Elsas, J. K. Jansson & J. T. Trevors (Eds.), *Modern Soil Microbiology (second edition)* (pp. 23-54). Boca Raton: CRC Press.

U.S. Environmental Protection Agency. (2007). *Ecological Soil Screening Levels for Pentachlorophenol*. Washington: Retrieved from http://rais.ornl.gov/documents/eco-ssl_pcp.pdf.

U.S. Environmental Protection Agency. (2013). ECOTOX User Guide: ECOTOXicology Database System. Version 4.0. http:/www.epa.gov/ecotox/

van den Brink, PJ, Brock, TCM, and Posthuma, L. (2001). The value of the species sensitivity distribution concept for predicting field effect: (Non-)confirmation of the concept using semifield experiments. In L. Posthuma, G. W. Suter II & T. P. Traas (Eds.), *Species sensitivity distributions in ecotoxicology* (pp. 155-193). Boca Raton: Lewis Publisher.

van der Heijden, MGA, Bardgett, RD, and van Straalen, NM. (2008). The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters, 11*(3), 296-310. doi: 10.1111/j.1461-0248.2007.01139.x

van Elsas, JD, Jansson, JK, and Trevors, JT. (2007). *Modern Soil Microbiology (second edition)*. Boca Raton: CRC Press.

van Straalen, NM, and Dennenman, CAJ. (1989). Ecotoxicological evaluation of soil quality criteria. *Ecotoxicology and Environmental Safety, 18*, 241-251.

van Straalen, NM, and van Leeuwen, CJ. (2001). European history of species sensitivity distributions. In L. Posthuma, G. W. Suter II & T. P. Traas (Eds.), *Species sensitivity distributions in ecotoxicology* (pp. 19-34). Boca Raton: Lewis Publisher.

Venturino, A, Montagna, CM, and de D'Angelo, AMP. (2007). Risk assessment of Magnacide® H herbicide at Rio Colorado irrigation channels (Argentina). Tier 3:Studies on native species. *Environmental Toxicology and Chemistry, 26*(1), 177.

Vig, K, Megharaj, M, Sethunathan, N, and Naidu, R. (2003). Bioavailability and toxicity of cadmium to microorganisms and their activities in soil: a review. *Advances in Environmental Research, 8*(1), 121-135. doi: http://dx.doi.org/10.1016/S1093-0191(02)00135-1

von der Ohe, PC, and Liess, M. (2004). Relative sensitivity distribution of aquatic invertebrates to organic and metal compounds. *Environmental Toxicology and Chemistry, 23*(1), 150-156.

Warne, MSJ, and van Dam, R. (2008). NOEC and LOEC Data Should No Longer Be Generated or Used. *Australasian Journal of Ecotoxicology, 14*(1), 1-5.

Weinstock, GM. (2012). Genomic approaches to studying the human microbiota. *Nature, 489*(7415), 250-256. doi: 10.1038/nature11553

Wheeler, JR, Grist, EPM, Leung, KMY, Morritt, D, and Crane, M. (2002). Species sensitivity distributions: data and model choice. *Marine Pollution Bulletin, 45*(1–12), 192-202. doi: http://dx.doi.org/10.1016/S0025-326X(01)00327-7

Wheeler, JR, Leung, KMY, Morritt, D, Sorokin, N, and Rogers, H. (2002). Freshwater to saltwater toxicity extrapolation using species sensitivity distributions. *Environmental Toxicology and Chemistry, 21*(11), 2459-5267.

Whittaker, RH. (1972). Evolution and Measurement of Species Diversity. *Taxon, 21*(2), 213-251.

Will, C, Thürmer, A, Wollherr, A, Nacke, H, Herold, N, Schrumpf, M, . . . Daniel, R. (2010). Horizon-Specific Bacterial Community Composition of German Grassland Soils, as Revealed by Pyrosequencing-Based Analysis of 16S rRNA Genes. *Applied and Environmental Microbiology, 76*(20), 6751-6759. doi: 10.1128/aem.01063-10

Winding, A, Hund-Rinke, K, and Rutgers, M. (2005). The use of microorgasnisms in ecological soil classification and assessment concepts. *Ecotoxicology and Environmental Safety, 62*(2), 230-248.

Wittebolle, L, Marzorati, M, Clement, L, Balloi, A, Daffonchio, D, Heylen, K, . . . Boon, N. (2009). Initial community evenness favours functionality under selective stress. *Nature, 458*(7238), 623-626. doi: 10.1038/nature07840

Xie, K, Xu, P, Yang, S, Tang, S, Zhang, F, Huang, X, . . . Zhang, H. (2011). The influence of paddy soil bacterial diversity affected by heavy metals contamination of Dabaoshan mine. *IEEE Xplore, 3*, 1705-1709. doi: 10.1109/iswrep.2011.5893576

Yankson, KK, and Steck, TR. (2009). Strategy for Extracting DNA from Clay Soil and Detecting a Specific Target Sequence via Selective Enrichment and Real-Time (Quantitative) PCR Amplification. *Applied and Environmental Microbiology, 75*(18), 6017-6021. doi: 10.1128/aem.00211-09

Yergeau, E, Lawrence, JR, Sanschagrin, S, Waiser, MJ, Korber, DR, and Greer, CW. (2012). Next-generation sequencing of microbial communities in the Athabasca River and its tributaries in

relation to oil sands mining activities. *Applied and Environmental Microbiology*. doi: 10.1128/aem.02036-12

Zielezny, Y, Groeneweg, J, Vereecken, H, and Tappe, W. (2006). Impact of sulfadiazine and chlorotetracycline on soil bacterial community structure and respiratory activity. *Soil Biology and Biochemistry, 38*(8), 2372-2380.

ANNEXE 1 : PROTOCOLES EXPÉRIMENTAUX

Protocole PE71A : Extraction de l'ADN total d'un échantillon de sol selon la méthode CTAB-DTT……..145

Protocole PE71B : Amplification de l'ADN par la méthode PCR…….…...……..160

Protocole PE71C : Migration DGGE (Denaturing Gradient Gel Electrophoresis) du produit PCR amplifié à partir de l'ADN extrait d'un sol……..…..…….170

Protocole PE71D : Utilisation de l'imageur Quantum ST4 et des logiciels de capture et d'analyse d'image Quantum-Capt et Bio-1D……...…….……184

Protocole PE71E : Contamination artificielle de sols au pentachlorophénol (PCP)……..……196

Protocole PE71F : Quantification du pentachlorophénol (PCP) par dosage par chromatographie en phase gazeuse couplée à un détecteur à ionisation de flamme……...……...…202

Protocole PE71G : Mesure du potentiel d'activité hydrolytique de la FDA dans un sol…..…….211

CIRAIG Centre Interuniversitaire de Référence sur l'Analyse, l'Interprétation et la Gestion du cycle de vie des produits, procédés et services Université de Montréal · École Polytechnique Montréal · HEC	**PROTOCOLE EXPÉRIMENTAL**
Protocole # :PE71A	Nombre de pages :15
Version :2	Date :12 janvier 2010
Auteur(s):Lucie Jean	
Approuvé par : Jonathan Lalande Louise Deschênes	Signatures : Date : ☐_____ _____ ☐_____ _____ ☐_____ _____
Titre : Extraction de l'ADN total d'un échantillon de sol selon la méthode CTAB-DTT	
Mots clés : Desoxyribonuclease acide, ADN, sol, extraction d'ADN, CTAB, DTT.	

OBJECTIF

La méthode d'extraction d'ADN vise à extraire l'ADN d'un échantillon de sol et à purifier afin d'éliminer toutes composantes organiques pouvant nuire à la quantification de l'ADN des micro-organismes se trouvant dans le sol à l'étude. Ce protocole vise donc à extraire l'ADN d'un sol selon la méthode de lyse cellulaire par bris mécanique des cellules microbiennes avec des billes de verre (zirconium-silice 0.1mm de diamètre) et d'une solution tampon de lyse composée d'hexadecyltrimethylammonium bromide- dithiothreitol (CTAB-DTT) (Bürgmann et al., 2001; Thakuria et al., 2008).

MATÉRIEL

Réactifs

- Hexadecyltrimethylammonium bromide (CTAB) (BioChemika Ultra, 52365, FW364.5) dessiccateur;
- Dithiothreitol (DTT) (Sigma, Grade molecular biology D9779, FW154.3) chambre froide 4°C;
- NaCl (Anachemia, ACS grade, AC-8304), FW58.44);
- EDTA (Fisher, molecular grade, BP120-500, FW372.24);
- Sodium Phosphate (NaH_2PO_4 .H_2O) (Anachemia, grade ACS);
- Acétate de potassium (Sigma, P1190, FW 98.14) dessiccateur;
- Acétate de sodium tri hydraté (Anachemia, 79948-380)
- Acide acétique (Fisher, UN-2789)
- TRIS (bio-Rad, 161-0119, FW121.14);
- Chloroforme (Fisher Scientific, C606-1);
- Alcool Isoamyl (3-méthyl-1-butanol) (Aldrich, 32002-1);
- Solution de Phenol :Chloroform :isoamyl alcohol (25 :24 :1) (Sigma,BioUltra, 77617) chambre froide 4°C;
- Polyethylene glycol 6000 (BioChemika Ultra, 81253) laboratoire;
- Phosphate de potassium (KH_2PO_4 et K_2HPO_4);
- Polyvynilpolypyrrolidone (Sigma, biology molecular grade, P6755) laboratoire;
- Solution HCl 10M;
- Solution HCl 3M;
- Solution NaOH 10M;
- Solution Hoechst 33258 à 2 mg/ml (Sigma, no 33258);
- Solution d'ADN 100µg/µl (D1501, Sigma)
- Eau déionisée

Équipements

- Billes zirconium-silice 0.1mm de diamètre autoclavables (Biospec product, no 11079101Z);

- Mini BeadBeater 8™ (Biospec Products, modèle 693, Bartlesville, Oklahoma);
- Fluoromètre (GENios, TECAN) couplé au logiciel Magellan (TECAN)
- Tubes microcentrifuges autoclavables 1,5 ml (Fisher Brand, no 05-669-30) et 2 ml (Fisher Scientific, no 05-407-25A);
- Tubes en polypropylène avec bouchon vissable 2 ml pour Mini BeadBeater (autoclavable);
- Centrifugeuse IEC (21000R, Needham Heights, USA);
- Micropipettes P1000, P200, P20 et P10ml (Gilson);
- Embouts à micropipettes stériles jaunes (Fisher Brand, no 21-197-8G) et bleus (Fisher Brand, no21-197-8F);
- Dispenseur à huit canaux (Nichiryo, modèle 8800);
- Microplaque à 96 puits;
- Vortexeur (Fisher Scientific, vortex, modèle génie 2, no 12-812, N.-Y., USA);
- pH-mètre (Accumet, modèle 25, pH/ion meter);
- Électrode (Orion Research Inc., Ross Sure-flow, Beverly, U.S.A.);
- Spatules;
- Béchers 500, 100 et 25 ml
- Ballons de 100 ml et 25 ml;
- Cylindre gradué de 100 ml;
- Pipette jaugée de 4 ml;
- Pipette graduée de 25 ml;
- Erlenmeyer sous vide de 2 l;
- Entonnoir Buchner et filtre whatman#1;
- Balance avec une précision de ± 0,0001 (Mettler Toledo, AB204, Switzerland);
- Tubes à centrifuge stériles jetables de 15 ml (Fisher Brand, no 05-539-5) et 50 ml (Fisher Brand, no 05-539-6);
- Papier aluminium;
- Bec brûleur;
- Pince;
- Bain-marie à 37 °C;

- Glace;
- Hotte chimique.

MÉTHODOLOGIE

Préparation des solutions

Solution tampon 0,2% hexadecyltrimethylammonium bromide (CTAB), 1mM dithioteitrol (DTT), 0.2M sodium phosphate buffer (pH 8), 0.1M NaCl, 50mM EDTA

1. Peser 0,2000 g de CTAB directement dans un bécher de 100 ml;
2. Peser 0,0154 g de DTT dans une coupelle et transférer dans le bécher de 100 ml;
3. Peser 2,7598 g de $NaH_2PO_4.H_2O$ dans une coupelle et le transférer dans le bécher de 100 ml;
4. Ajouter environ 80 ml H_2O, ajouter un barreau magnétique et agiter sur une plaque magnétique;
5. Ajuster le pH à 8 à l'aide du pH-mètre avec une solution de NaOH 10 M (Voir le PSO E3 pour l'utilisation du pH-mètre.);
6. Peser 0,5844 g de NaCl dans une coupelle et transférer dans le bécher de 100 ml;
7. Peser 1,8612 g de EDTA dans une coupelle et transférer dans le bécher de 100 ml;
8. Agiter sur une plaque magnétique jusqu'à dissolution complète des solides;
9. Transférer quantitativement dans un ballon de 100 ml et compléter jusqu'au trait jauge avec de l'eau distillée.

Solution d'acétate de potassium 5M

1. Peser 12,2675 g directement dans un bécher de 25 ml;
2. Ajouter environ 20 ml H2O, ajouter un barreau magnétique et agiter sur une plaque magnétique jusqu'à dissolution complète;
3. Transférer quantitativement dans un ballon de 25 ml et compléter jusqu'au trait jauge avec de l'eau distillée.

Solution de Polyethylene glycol 6000 (PEG) 20%, 2.5M NaCl

1. Peser 5,0000 g PEG directement dans un bécher de 25 ml;
2. Peser 3,6525 g de NaCl dans une coupelle et transférer dans le bécher de 25 ml;
3. Ajouter environ 20 ml H2O, ajouter un barreau magnétique et agiter sur une plaque magnétique jusqu'à dissolution complète;
4. Transférer quantitativement dans un ballon de 25 ml et compléter jusqu'au trait jauge avec de l'eau distillée.

Solution tampon TE (50mM TRIS-HCl, 10mM EDTA, pH8)

1. Peser 0,6057 g de TRIS et le transférer dans un bécher de 100 ml;
2. Ajouter environ 80 ml H2O, ajouter un barreau magnétique et agiter sur une plaque magnétique;
3. Ajuster le pH à 8 à l'aide du pH-mètre avec une solution de HCl 10 M (Voir le PSO E3 pour l'utilisation du pH-mètre.);
4. Peser 0,3722 g d'EDTA dans une coupelle et transférer dans le bécher de 100 ml et agiter jusqu'à dissolution complète des solides;
5. Transférer quantitativement dans un ballon de 100 ml et compléter jusqu'au trait jauge avec de l'eau distillée.

Solution Chloroform :isoamyl alcohol (24 :1)

1. Travailler sous la hotte chimique pour la préparation de cette solution
2. À l'aide d'une pipette jaugée 4 ml, mettre 4 ml d'alcool isoamyl dans un ballon de 100 ml;
3. À l'aide d'un cylindre gradué 100 ml, ajouter le chloroforme dans le ballon de 100 ml et compléter jusqu'au trait de jauge;
4. Bouchonner le ballon de 100 ml et l'inverser à treize reprises.

Solution Phenol :Chloroform :isoamyl alcohol (25 :24 :1 pH6.8)

1. Travailler sous la hotte chimique pour la préparation de cette solution
2. Prélever 25 ml de la solution Phenol :Chloroform :isoamyl alcohol (25 :24 :1) et transférer dans un bécher de 25 ml;
3. Ajuster le pH à 6.8 à l'aide du pH-mètre avec une solution de HCl 10 M (Voir le PSO E3 pour l'utilisation du pH-mètre.).

Solution de phosphate de potassium (20 mM, pH7.4)

1. Peser 1,3600 g de KH2PO4 et le transférer dans un bécher de 500 ml;
2. Ajouter environ 80 ml H2O, ajouter un barreau magnétique et agiter sur une plaque magnétique;
3. Transférer quantitativement dans un ballon de 500 ml et compléter jusqu'au trait jauge avec de l'eau distillée;
4. Peser 5,4400 g de K2HPO4 et le transférer dans un bécher de 2 l;
5. Ajouter environ 1600 ml H2O, ajouter un barreau magnétique et agiter sur une plaque magnétique;
6. Transférer quantitativement dans un ballon de 2 l et compléter jusqu'au trait jauge avec de l'eau distillée;
7. Transférer dans un bécher de 2 l environ 250 ml de solution KH2PO4 et ajuster le pH à 7,4 (voir PSO E3) avec la solution de K2HPO4. Pour 500 ml de KH2PO4, il faudra normalement un peu moins de 2 l de K2HPO4 afin d'arriver à un pH de 7,4. Lorsque ce pH est atteint, verser la solution obtenue dans une bouteille en verre de 2 l. Reprendre jusqu'à ce que toute la solution de KH2PO4 ait été utilisée.

Solution d'acétate de sodium (3M, pH5.2)

1. Peser 40,8240 g d'acétate de sodium et le transférer dans un bécher de 100 ml;
2. Ajouter environ 60 ml H2O, ajouter un barreau magnétique et agiter sur une plaque magnétique;
3. Ajuster le pH à 5.2 à l'aide du pH-mètre avec une solution d'acide acétique concentré

(Voir le PSO E3 pour l'utilisation du pH-mètre.);

4. Transférer quantitativement dans un ballon de 100 ml et compléter jusqu'au trait jauge avec de l'eau distillée.

Préparation de 20 ml d'une solution Hoechst (0.8 µg/ml)

Il faut d'abord préparer une solution Hoechst de 2mg/ml

1. Peser 2mg de colorant Hoechst dans un tube eppendorf de 1,5ml stérile.
2. Avec une micropipette, ajouter 1ml de tampon TE dans le tube.
3. Mélanger pour dissoudre le colorant Hoechst.
4. Envelopper le tube dans du papier d'aluminium et conserver à 4°C jusqu'à l'utilisation.

Dilutions de la solution Hoechst (2mg/ml) (préparées le jour même de l'analyse par fluorométrie). On effectue une dilution sérielle à partir de trois tubes : 2 tubes de 15 ml et 1 tube de 50 ml. Les quantités peuvent être modifiées selon le nombre d'échantillons à analyser par fluorométrie. ***Cependant, s'assurer de conserver les mêmes proportions.***

1. À l'aide de la micropipette P200, prélever 25 µl de la solution Hoechst concentrée (2 mg/ml) et le transférer dans un tube à centrifugeuse jetable de 15 ml;
2. À l'aide de la pipette automatique P5000, ajouter 2.5 ml de tampon TE;
3. Boucher le tube et vortexer 10 secondes;
4. À l'aide de la pipette automatique P5000, prélever 1 ml de la solution vortexée et la transférer dans un nouveau tube à centrifugeuse jetable de 15 ml;
5. À l'aide de la pipette automatique P5000, lui ajouter 4 ml de tampon TE;
6. Boucher le tube et vortexer de nouveau pendant 10 secondes;
7. À l'aide de la pipette P5000, prélever 4 ml de la solution obtenue à l'étape 5 et la transférer dans un nouveau tube à centrifugeuse jetable de 50 ml;
8. À l'aide d'une pipette graduée de 25 ml, lui ajouter 16 ml de tampon TE;
9. Boucher et recouvrir le tube de 50 ml contenant la solution Hoechst (0.8 µg/ml) d'un papier aluminium, car cette solution réagit à la lumière et perd sa fluorescence.

Préparation des colonnes de PVPP

Solution de PVPP

1. Peser 20,00 g de PVPP et le transférer dans un bécher de 500 ml;
2. Ajouter 100 ml de HCL 3M;
3. Filtrer cette solution à l'aide d'un entonnoir Buchner avec papier filtre Whatman#1;
4. Laver le PVPP avec 500 ml d'eau déionisée;
5. Laver de nouveau avec 500 ml de tampon KPO_4 (20 mM, pH 7.4);
6. Transférer le PVPP dans une bouteille propre et laisser sécher.

Préparation des colonnes de PVPP

1. Transférer le PVPP dans un bécher de 100 ml et ajouter 50 ml de tampon KPO_4 (20 mM, pH 7.4);
2. Introduire environ 1 mm de laine dans une seringue Hamilton de 1 ml;
3. Dans un tube à centrifuge de 15 ml, introduire une tube Eppendorf de 1,5 ml, puis la seringue de 1 ml (Figure A1-1);
4. À l'aide d'une seringue de 10 ml, prélever la solution de PVPP;
5. Transférer 1 ml de cette solution dans la seringue de 1 ml du montage;
6. Centrifuger à 1000 rpm pendant 1 min

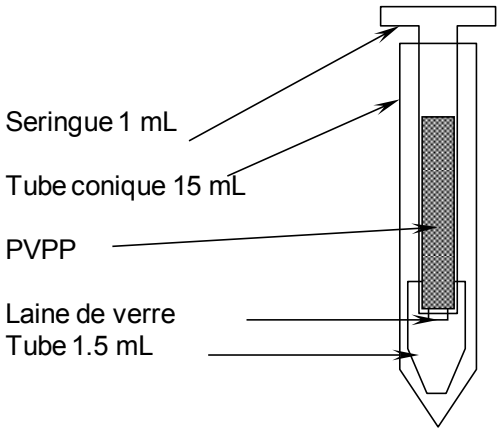

Figure A1-1 : Montage d'une colonne de PVPP

Procédure d'extraction de l'ADN d'un sol (toujours déposer les échantillons sur de la glace lors des temps d'attente afin de limiter la dégradation de l'ADN)

1. Peser 0,3750 g de billes Biospec de 0.1 mm de diamètre dans un tube adapté de 2 ml pour la machine Mini BeadBeater et bouchonner;
2. Autoclaver ce tube pendant 15 minutes à 120°C en position "SOLIDS";

Travailler en mode aseptique (le rayon d'une flamme suffit)

3. Peser directement 0,3000 g de sol dans le tube adapté pour la machine Mini BeadBeater contenant 0,3750 g de billes Biospec de 0.1 mm de diamètre préalablement autoclavées. (Voir les recommandations pour les différents types de sol);
4. À l'aide d'une pipette automatique P1000, ajouter 650 µl de tampon CTAB-DTT 0,2% dans le tube contenant le sol et les billes et bouchonner;
5. Introduire le tube dans la machine Mini BeadBeater et homogénéiser à vitesse maximale pendant 90 secondes;
6. Centrifuger le tube adapté Mini BeadBeater à 16 000 g pendant 5 min à 4°C;
7. Récolter le surnageant, environ 500 µl, à l'aide d'une pipette automatique P1000 et le transférer dans la colonne de PVPP préparé au point 3.2.2;
8. Centrifuger à 1000 rpm suffisamment longtemps pour que tout l'échantillon passe au travers de la colonne, cela devrait prendre entre 6 et 8 minutes;
9. Récolter les échantillons à l'aide d'une pipette automatique P1000 et mettre ces derniers dans des tubes microcentrifuge de 2 ml préalablement autoclavés (le volume de l'extrait devrait se situer entre 550 et 600 µl, le PVPP ayant relâché un petit volume de phosphate de potassium dans l'extrait);
10. À l'aide d'une pipette automatique P200, ajouter 65 µl d'acétate de potassium 5M. Attention la concentration finale doit être de 0,5M d'acétate de potassium (dilution 1 :10);
11. Incuber le tube sur glace pendant 5 min;
12. Centrifuger à 16 000 g pendant 30 min à 4°C;
13. Récolter le surnageant, entre 615 et 665 µl à l'aide d'une pipette automatique P1000 et le

transférer dans un tube microcentrifuge de 2 ml préalablement autoclavé;

14. À l'aide d'une pipette automatique P1000, ajouter 650 µl de la solution Polyethylene glycol 6000 (PEG) 20%, 2.5M NaCl (ratio d'environ 1:1);
15. Incuber le tube dans un bain tempéré à 37°C pendant 1 heure;
16. Centrifuger à 16 000 g pendant 30 min à 20°C;
17. Retirer le surnageant en inversant le tube, le culot contient l'ADN;
18. Dissoudre le culot (faire plusieurs mouvements de va et vient avec la micropipette P1000) avec 400 µl de tampon TE;

Travailler sous la hotte chimique pour les étapes qui suivent

1. Ajouter 400 µl de la solution Phenol :Chloroform :isoamyl alcohol (25 :24 :1) et vortexer quelques secondes;
2. Centrifuger à 16000 g pendant 5 min à 4°C;
3. Récolter 350 µl de la phase aqueuse (phase supérieure) à l'aide d'une pipette automatique P1000 et le transférer dans un tube microcentrifuge de 2 ml préalablement autoclavé (procéder délicatement afin de ne récolter que la phase aqueuse);
4. Ajouter 350 µl de la solution Chloroform :isoamyl alcohol (24 :1) (ratio de 1:1) et inverser doucement quelques secondes;
5. Centrifuger à 16000 g pendant 1 min à 4°C;
6. Répéter les étapes 21, 22, 23 et de nouveau 21 (en récoltant cette fois 325 µl de la phase aqueuse);
7. À l'aide d'une pipette automatique P20, ajouter 32,5 µl de la solution d'acétate de sodium pH5.2 3M (1/10 volume);
8. À l'aide d'une pipette automatique P1000, ajouter 715 µl d'éthanol 100% (2 volumes);
9. Incuber le tube à -20°C pendant 15 min;
10. Centrifuger à 16 000 g pendant 10 min à 4°C;
11. Retirer le surnageant en inversant le tube, le culot contient l'ADN;

12. Avec une micropipette P1000, ajouter 500µl d'éthanol 70% dans chacun des échantillons d'ADN;
13. Centrifuger les tubes à 21000g pendant 2 minutes à 4°C;
14. Disposer du surnageant dans un petit bécher;
15. Assécher complètement les échantillons d'ADN à l'air ambiant.

Il faut s'assurer que l'éthanol est complètement évaporé avant de poursuivre, sinon il sera impossible de suspendre l'ADN (il restera sous forme de précipité). À cette étape-ci, il est possible de laisser l'éthanol s'évaporer durant la nuit en laissant les échantillons sous la hotte biologique.

16. Dissoudre le culot (faire plusieurs mouvements de va et vient avec la micropipette P200) avec 100 µl de tampon TE.

Mesure de la concentration d'ADN dans les extractions

On établit la courbe standard d'ADN à partir de concentrations connues d'ADN (voir l'exemple d'une courbe standard en annexe)

Préparation de la solution mère d'ADN

1. Dans un tube à centrifugation de 15ml, peser 1mg d'ADN;
2. Avec une micropipette de 10ml, ajouter 10ml de tampon TE;
3. On obtient une solution mère d'ADN de 100µg/ml;
4. Laisser l'ADN se solubiliser pendant 24 heures dans une chambre froide à 4°C.

Préparation des standards d'ADN

À partir d'un standard d'ADN de 100 µg/ml, préparer des standards pour les concentrations suivantes : 25 ug/ml, 20 µg/ml, 15 µg/ml, 10 µg/ml, 5 µg/ml et 2,5 µg/ml;

Préparer le standard de 25 µg/ml :

1. À l'aide d'une seringue de 250µl, prélever 250µl de la solution d'ADN de 100µg/ml et le transférer dans un tube microcentrifuge stérile de 1,5 ml;
2. À l'aide d'une seringue de 1ml, y ajouter 750µl de tampon TE;

Préparer le standard de 20 µg/ml :

1. À l'aide d'une seringue de 250µl, prélever 200µl de la solution d'ADN de 100µg/ml et le transférer dans un tube microcentrifuge stérile de 1,5 ml;
2. À l'aide d'une seringue de 1ml, y ajouter 800µl de tampon TE;

Préparer le standard de 15 µg/ml :

1. À l'aide d'une seringue de 250µl, prélever 150µl de la solution d'ADN 100µg/ml et le transférer dans un tube microcentrifuge stérile de 1,5 ml;
2. À l'aide d'une seringue de 1000µl, y ajouter 850µl de tampon TE;

Préparer le standard de 10 µg/ml :

1. À l'aide d'une seringue de 100µl, prélever 100µl de la solution d'ADN 100µg/ml et le transférer dans un tube microcentrifuge stérile de 1,5 ml;
2. À l'aide d'une seringue de 1ml, y ajouter 900µl de tampon TE;

Préparer le standard de 5 µg/ml :

1. À l'aide d'une seringue de 50 µl, prélever 50µl de la solution d'ADN 100 µg/ml et le transférer dans un tube microcentrifuge stérile de 1,5 ml;
2. À l'aide d'une seringue de 1ml, y ajouter 950 µl de tampon TE;

Préparer le standard de 2,5 µg/ml :

1. À l'aide d'une seringue de 25 µl, prélever 25 µl de la solution d'ADN 100 µg/ml et le transférer dans un tube microcentrifuge stérile de 1,5 ml;
2. À l'aide d'une seringue de 1ml, y ajouter 975 µl de tampon TE;

Vortexer les solutions standards et les déposer sur de la glace jusqu'à leur utilisation. Conserver les standards au congélateur à –20°C après leur utilisation.

Quantification de la concentration en ADN dans les extractions

Lecture au fluoromètre

On procède à l'analyse de l'ADN par fluorométrie en utilisant une microplaque de 96 puits dans laquelle on introduit des quantités fixes d'ADN (50 µl) et de solution Hoechst DE 0,8 µg/ml (250 µl). On dépose ensuite la microplaque dans le lecteur à microplaque. On configure les paramètres du logiciel SPF. Les paramètres à modifier sont la longueur d'onde d'excitation (360 nm) et la longueur d'onde d'émission (460 nm). On doit également mettre le brassage de la microplaque à ON.

1. À l'aide d'une micropipette P200, prélever à deux reprises (duplicata) 50 µl de tampon TE et déposer l'échantillon dans les puits (section A1, A2) de la microplaque. Ces échantillons serviront de blanc pour définir la courbe standard;
2. À l'aide d'une micropipette P200, prélever à deux reprises (duplicata) 50 µl de la solution standard d'ADN 2,5 µg/ml et déposer l'échantillon dans les puits (section B1, B2) de la microplaque;
3. Répéter l'étape précédente pour toutes les solutions standards d'ADN en les déposant dans les puits C1, C2 jusqu'à G1, G2;
4. À l'aide de la micropipette P200, prélever 50 µl de l'échantillon d'ADN et le déposer dans le puits A3 de la microplaque;
5. À l'aide d'une pipette automatique p1000, ajouter 250 µl de solution Hoechst 0.8 µg/ml dans les puits contenant 50 µl d'ADN. La microplaque est lue directement par le lecteur à microplaque (voir le protocole PSO-E19 pour le fonctionnement du fluoromètre);
6. Convertir les valeurs de fluorescence obtenues à partir de la courbe standard d'ADN.

Vérification de l'ADN extrait par migration sur gel d'agarose 1%

1. Peser directement dans un erlenmeyer de 250 ml 0,5000 g d'agarose;
2. Prélever dans un cylindre gradué 50 ml de tampon TAE 1X et verser dans l'erlenmeyer contenant l'agarose;
3. Faire chauffer cette solution afin de solubiliser complètement l'agarose (mélanger de temps à autre pendant le chauffage), le chauffage devrait être d'environ 1 minute

4. Verser cette solution dans un moule à gel en plastique dont les deux extrémités ont été « bouchés » par du ruban autocollant (masking tape);
5. Fixer rapidement le peigne à l'extrémité du gel avant que ce dernier ne se solidifie;
6. Laisser le gel se solidifier (entre 30 et 45 minutes) en s'assurant que le moule à gel est de niveau. Lorsque le gel s'est solidifié, retirer délicatement le peigne et déposer le moule dans le bain de migration rempli avec du tampon TAE 1X;
7. Déposer dans des eppendorf (de 0,2 ml) 1 µl de tampon de migration à l'aide d'une pipette automatique de 2 µl;
8. Dans le premier eppendorf, déposer 5 µl de la solution standard d'ADN (Ladder 1 kb) à l'aide d'une pipette automatique de 20 µl;
9. Dans les autres eppendorf, déposer 5 µl des échantillons d'ADN à analyser à l'aide d'une pipette automatique de 20 µl;
10. Injecter la totalité du contenu des eppendorf dans les puits formés dans le gel à l'aide d'une pipette automatique de 20 µl (attention, il faut réaliser ces manipulations très délicatement);
11. Brancher l'alimentation du bain à migration et régler la différence de potentiel à 100 V;
12. Laisser migrer l'ADN dans le bain durant environ 45 à 60 minutes;
13. Couper le courant de l'appareil et débrancher le fil d'alimentation du bain à migration;
14. Déposer le gel dans un bain de bromure d'éthydium (coloration environ 10 minutes et décoloration dans de l'eau déionisée durant 15 minutes) ou de SYBR green (coloration environ 30 minutes, pas de décoloration);
15. Visualiser le gel à l'aide de l'appareil Quantum ST4 (voir PE71D).

SANTÉ ET SECURITÉ

Le port des gants est exigé pendant toutes les manipulations pour l'extraction d'ADN. La manipulation des solvants s'effectue en tout temps sous la ventilation d'une hotte chimique.

Consulter les fiches signalétiques du phénol, chloroforme et de l'alcool isoamyl avant leur utilisation.

RECOMMANDATION(S)

Le type de sol est un paramètre important lors de l'extraction de l'ADN. La quantité de sol à analyser doit être évaluée expérimentalement pour chaque type de sol utilisé. On peut faire varier la quantité de sol de 100 à 500 mg seulement.

CIRAIG Centre Interuniversitaire de Référence sur l'Analyse, l'Interprétation et la Gestion du cycle de vie des produits, procédés et services Université de Montréal · École Polytechnique Montréal · HEC	**PROTOCOLE EXPÉRIMENTAL**
Protocole # : PE71B	Nombres de pages :10
Version :2	Date :01/04/12
Auteur(s): Jonathan Lalande	
Approuvé par : Lucie Jean	Signatures : Date : ☒_____ _____ ☐_____ _____ ☐_____ _____
Titre : Amplification de l'ADN par la méthode PCR	
Mots clés : ADN, sol, PCR, bactéries, champignons, électrophorèse	

OBJECTIF

L'objectif de la méthode PCR vise à amplifier une région spécifique de l'ADN de divers organismes. La région de l'ADN et les organismes ciblés dépendent des amorces utilisées afin de procéder à l'amplification, qui peuvent être très spécifiques ou quasi-universelles pour un type d'organisme (ex. bactéries). L'amplification PCR doit permettre d'obtenir une quantité suffisante d'ADN afin d'analyser le produit par séquençage ou sur un gel électrophorèse. Pour certaines applications (séquençage), le produit amplifié devra être purifié avant de procéder à l'analyse.

Bien qu'un protocole général soit présenté ici, la concentration des différents réactifs ainsi que les

conditions de l'amplification (dilution des extraits d'ADN, nombre de cycles, températures, temps…) dépend de la pureté des extraits d'ADN, du type de sol étudié, des amorces utilisées, du type d'enzyme employé ainsi que de la région à amplifier (%G+C). Ces conditions doivent être optimisées.

MATÉRIEL

- Tubes microcentrifuges eppendorf 0.5 et 1,5 ml (Fisher Scientific);
- Embouts pour pipette stériles;
- Seringue de 1 ml;
- Filtre millipore stérile de 0.22 µm;
- Support à tubes microcentrifuges;
- Pipettes P2, P20, P200 et P1000;
- H_2O stérile;
- Solution BSA, bovine serum albumin (Sigma, cat A-2153);
- Solution d'amorces (10pmoles/µl, Alpha DNA);
- Phusion® High-Fidelity PCR Master Mix (Thermo Scientific, cat no F-532L);
- Solution de $MgCl_2$ (100mM);
- Agarose (Invitrogen, cat no 15510-027);
- Bromure d'éthidium 10 mg/ml (Sigma, cat no. E1510-10ml);

ATTENTION, SUBSTANCE HAUTEMENT TOXIQUE

- EZ Load™ Precision Molecular Mass Standard (Bio-Rad, cat no. 170-8356);
- Solution tampon TAE 50X (Tris, acide acétique, EDTA);
- Appareil électrophorèse (Power Pac Basic™, 300V, 400mA, 75W, BIO-RAD);
- Support électrophorèse (MINI SUB® CELL GT, BIO-RAD);
- Appareil PCR (Techne, Genius, FGENO2TP).
- Kit d'extraction de gel « GenElute™ Gel Extraction kit » (Sigma, cat. NA1111) (optionnel);

- Éthanol 100%;
- Spatule métallique ou scalpel (optionnel).

MÉTHODOLOGIE

Préparation de la solution BSA

1. À la balance, peser directement dans un tube eppendorf 1.5 ml, 20 mg de BSA;
2. À l'aide d'une pipette P1000, ajouter 1 ml d'eau déionisée et agiter en effectuant un mouvement de va et vient avec la pipette;
3. À l'aide d'une seringue de 1 ml et d'un filtre millipore 0.22 µm, filtrer la solution de façon stérile dans un tube eppendorf 1.5 ml stérile;
4. Faire des aliquots dans des tubes eppendorf stériles de 0.5 ml. Le volume des aliquots dépend du volume généralement utilisé pour une série d'amplification (nombre d'échantillons X volume ajouté par échantillion);
5. Congeler à –20°C jusqu'à utilisation.

Préparation des solutions d'amorces (10 pmol/µl)

1. À la réception des amorces, dissoudre ces dernières dans de l'eau déionisée stérile de façon à obtenir des solutions dont la concentration est de 100 pmol/µl;
2. À l'aide d'une pipette P1000, prélever 900 µl d'eau déionisée stérile et transférer dans un tube eppendorf stérile de 1.5 ml;
3. À l'aide d'une pipette P200, prélever 100 µl de la solution d'amorce (100 pmoles/µl) et la transférer dans le tube eppendorf stérile de 1.5 ml;
4. Faire des aliquots dans des tubes eppendorf stériles de 0.5 ml (le volume des aliquots à préparer dépend de la quantité d'amorces ajoutée lors des réactions PCR et du nombre d'échantillons à analyser simultanément);
5. Ces étapes doivent être répétées pour chacune des amorces.

Préparation d'une solution TAE 50X

1. À la balance, peser 121g de TRIS dans une coupelle et transférer dans un bécher de 500 ml;

2. Sous une hotte chimique, prélever 28.55 ml d'acide acétique glacial avec une pipette graduée de 50 ml;
3. Ajouter l'acide acétique glacial dans le bécher en prenant soin de rincer la coupelle ayant servi à peser le TRIS;
4. À l'aide d'une pipette graduée de 50 ml, prélever 50 ml de la solution d'EDTA.2Na 0.5M pH 8.0 préparée au point 3.1.2 du protocole d'extraction d'ADN (PE42B) et transférer dans le bécher de 500 ml;
5. Ajouter de l'eau déionisée jusqu'à ce que le volume dans le bécher soit d'environ 400ml;
6. Insérer un barreau magnétique et agiter jusqu'à dissolution complète du TRIS;
7. Transvider le contenu du bécher dans un ballon jaugé de 500ml;
8. Compléter à 500 ml avec H2O déionisée en prenant soin de rincer au moins trois fois le bécher ayant contenu la solution;
9. Sceller le ballon et l'inverser au moins treize fois pour homogénéiser la solution;
10. Répartir le contenu du ballon dans deux bouteilles en verre de 500ml avec bouchon en téflon;
11. Autoclaver les bouteilles pendant 30minutes.

Préparation de la solution TAE 1X

1. À l'aide d'une micropipette P10 ml, prélever 10 ml de la solution TAE 50X et transférer dans une autre bouteille de 500 ml;
2. À l'aide d'un cylindre gradué de 500 ml, prélever 490 ml d'eau déionisée et la transférer dans cette autre bouteille de 500 ml. On obtient donc une solution de TAE 1X.

Amplification PCR

La recette à utiliser dépend de ce que l'on cherche à amplifier. Cette dernière doit être optimisée pour chaque nouveau type de sol ou paire d'amorces utilisée. Certains additifs peuvent être ajoutés afin d'optimiser l'amplification (voir le guide de dépannage fourni avec les tubes d'enzymes Finnzymes ou différent guides disponibles en ligne). Les recettes utilisées avec certaines paires d'amorces sont données en annexes.

Une seule amplification

1. Préparer, dans un epperdorf stérile de 200 ul, la recette unitaire présenté au Tableau A1-1. Ajouter les réactifs dans l'ordre du tableau;
2. Ajouter 1µl de solution d'ADN (5-10 ng/µl) issue d'une dilution de l'extrait d'ADN;

La concentration peut varier. Réaliser des dilutions de l'extrait d'ADN de façon à ce que l'amplification PCR ne soit pas inhibée. Lorsque les extractions sont réalisées à l'aide du protocole PE71B, une dilution 1:2 est normalement suffisante. Certains additifs peuvent également aider à prévenir l'inhibition (BSA, $MgCl_2$, Glycerol, DMSO…).

Tableau A1-1: Recette PCR pour les amorces 341f-GC et 534r (DGGE).

Réactif	**Volume (µl)**
Phusion PCR master mix	25
Eau déionisée stérile	18
BSA	3
$MgCl_2$	1
Amorce 341f-GC (10 pmoles/µl)	1
Amorce 534r (10 pmoles/µl)	1

3. Insérer le tube dans l'appareil Techne Genius et allumer l'appareil;
4. S'assurer que le programme 1 a été sélectionné;
5. Appuyer sur le bouton START;
6. Fermer le couvercle et laisser le programme rouler.

Plusieurs amplifications

Lorsqu'on a plus d'une amplification à faire, multiplier les volumes présentés au Tableau A1-1 par le nombre d'amplifications à réaliser. Tenir compte qu'il faut au moins un blanc de méthode (considérer donc au moins une réaction de plus par rapport au nombre échantillons à analyser). En règle générale, il faut préparer une réaction de plus que le nombre d'échantillons à analyser, en incluant le blanc de méthode;

1. Préparer la recette PCR dans un tube eppendorf stérile de 1,5ml. Ajouter les ingrédients dans le même ordre que décrit dans les tableaux disponibles en annexes;
2. Prélever 49 µl de cette solution et la distribuer dans les tubes eppendorf stériles de 0,5ml qui seront utilisés pour les échantillons d'ADN à tester;
3. Avec une pipette P2, ajouter 1 µl des échantillons d'ADN dans chacun des tubes;
4. Insérer les tubes dans l'appareil Techne Genius et allumer l'appareil;
5. S'assurer que le programme 1 a été sélectionné;
6. Appuyer sur le bouton START;
7. Fermer le couvercle et laisser le programme rouler.

Amplification

Les conditions des cycles d'amplification doivent être optimisées en fonction du T_m des amorces et de la recette PRC utilisée. Le touchdown PCR vise à augmenter la spécificité du produit amplifié. Le détail des cycles utilisés pour certaines paires d'amorces sont fournies en annexes. Les conditions utilisées pour les amorces 341f-GC et 534r sont présentées dans le Tableau A1-2.

Tableau A1-2: Détail de l'amplification pour les amorces 341f-GCet 534r.

Étape	Détail
Dénaturation initiale	94°C durant 90 secondes
Touchdown (10 cycles)	Dénaturation : 94°C, 10 secondes Appariement : 67.5°C à 62.5°C (-0.5°C/cycle), 10 secondes Élongation : 72°C, 10 secondes
Amplification (20-30 cycles)	Dénaturation : 94°C, 10 secondes Appariement : 62.5°C, 10 secondes Élongation : 72°C, 10 secondes
Élongation finale	72°C, 5 minutes
Hold	4°C, indéfiniment

Vérification du produit PCR sur un gel d'agarose 2.0 % (entre 1.5% et 2%)

1. À la balance, peser dans une coupelle 0.8 g d'agarose et transférer dans un erlenmeyer de 250 ml;
2. À l'aide d'un cylindre gradué de 50 ml, mesurer un volume de 40 ml de la solution TAE 1X et l'ajouter à l'erlenmeyer de 250 ml;
3. Faire chauffer la solution dans une micro-onde d'abord pendant environ 30 secondes. Remuer légèrement l'erlenmeyer. Remettre dans la micro-onde durant 20 secondes. Remuer légèrement l'erlenmeyer. Remettre dans la micro-onde durant 10 secondes;
4. Laisser refroidir légèrement la solution et y ajouter 0.75µl d'une solution de bromure d'éthidium 10 mg/ml;
5. Couler la solution chaude dans le support à migration électrophorèse (on doit mettre du ruban adhésif aux extrémités du support pour maintenir la solution dans le support voir Figure A1-2);
6. Fixer le peigne approprié sur le support (puits étroits);
7. Couvrir pour protéger de la lumière et laisser gélifier pendant 30 min à la température ambiante;
8. Prendre un nombre X (nombre d'échantillons à analyser) de tubes eppendorf stériles de 0,5 ml. À l'aide d'une pipette P2, y déposer 1 µl de tampon de migration (colorant bleu) accompagnant le EZ load;
9. À ces tubes, ajouter 5µl de solution d'ADN (produit PCR) à analyser;
10. Retirer le peigne et le ruban adhésif du support et déposer le support dans la cuve électrophorèse, remplie de solution TAE 1X;
11. Avec la pipette P20, charger les puits avec les solutions d'ADN et les marqueurs. S'assurer de bien noter les correspondances puits-échantillon;
12. Avec une pipette P20, injecter 5µl de la solution contenant des standards massiques d'ADN (EZ load mass standard) dans un puits;
13. Mettre le couvercle en place en prenant soin de bien brancher les pôles au bon endroit (rouge

avec rouge et noir avec noir);

14. Couvrir pour protéger de la lumière et allumer l'appareil;
15. Mettre l'appareil en mode mA (appuyer sur la touche permettant de sélectionner le mode jusqu'à ce que la lumière soit vis-à-vis le mode mA);
16. Appuyer sur le bouton STOP, jusqu'à ce que le symbole dEq apparaisse;
17. Mettre l'appareil en mode V;
18. S'assurer que le voltage est de 100 à l'écran. Si ce n'est pas le cas, ajuster avec les flèches situées à droite de l'écran;
19. Mettre l'alimentation en marche (appuyer sur RUN) et laisser migrer pendant 45 minutes;
20. Débrancher l'appareil électrophorèse, retirer le couvercle et le support;
21. Analyser le gel sous une lampe UV et prendre une photo (voir PE71D pour le fonctionnement de l'appareil Quantum ST4).

Excision d'une bande et purification du produit amplifié (si nécessaire)

1. À la balance, peser dans une coupelle 1.2g d'agarose et transférer dans un erlenmeyer de 250 ml;
2. À l'aide d'un cylindre gradué de 100 ml, mesurer un volume de 60 ml de la solution TAE 1X et l'ajouter à l'erlenmeyer de 250 ml;
3. Faire chauffer la solution dans une micro-onde d'abord pendant environ 30 secondes. Remuer légèrement l'erlenmeyer. Remettre dans la micro-onde durant 20 secondes. Remuer légèrement l'erlenmeyer. Remettre dans la micro-onde durant 10 secondes;
4. Laisser refroidir légèrement la solution et y ajouter 0.5µl d'une solution de bromure d'éthidium 10 mg/ml;
5. Couler la solution chaude dans le support à migration électrophorèse (on doit mettre du ruban adhésif aux extrémités du support pour maintenir la solution dans le support voir Figure A1-2);
6. Fixer le peigne approprié sur le support (puits larges);
7. Couvrir pour protéger de la lumière et laisser gélifier pendant 30 min à la température

ambiante;

8. Ajouter aux tubes eppendorf contenant les amplifications (il devrait y avoir entre 45 et 50µl de solution) 10µl de tampon de migration (2X gel loading dye);
9. Retirer le peigne et le ruban adhésif du support et déposer le support dans la cuve électrophorèse, remplie de solution TAE 1X;
10. Avec la pipette P200, charger les puits avec les solutions d'ADN et les marqueurs (attention aux bulles d'air). S'assurer de bien noter les correspondances puits-échantillon;
11. Avec une pipette P20, injecter 5µl de la solution contenant des standards massiques d'ADN (EZ load mass standard) dans un puits;
12. Mettre le couvercle en place en prenant soin de bien brancher les pôles au bon endroit (rouge avec rouge et noir avec noir);
13. Couvrir pour protéger de la lumière et allumer l'appareil;
14. Mettre l'appareil en mode mA (appuyer sur la touche permettant de sélectionner le mode jusqu'à ce que la lumière soit vis-à-vis le mode mA);
15. Appuyer sur le bouton STOP, jusqu'à ce que le symbole dEq apparaisse;
16. Mettre l'appareil en mode V;
17. S'assurer que le voltage est de 100 à l'écran. Si ce n'est pas le cas, ajuster avec les flèches situées à droite de l'écran;
18. Mettre l'alimentation en marche (appuyer sur RUN) et laisser migrer pendant 45 minutes, ou jusqu'à ce que le marqueur de migration ayant la vitesse de migration la plus rapide atteigne minimalement le centre du gel;
19. Débrancher l'appareil électrophorèse, retirer le couvercle et le support;
20. Déposer le gel sur le transilluminateur du système Quantum ST4;
21. Mettre des lunettes de protection contre les rayons UV, allumer l'appareil (enfoncer les boutons « Power », « UV » ainsi que le bouton marqué D'un triangle contenant un point d'exclamation (ce bouton fait en sorte que le rayonnement UV continue lorsque la porte de l'appareil est ouverte)); les bandes d'ADN devraient apparaître en rouge sur le gel;

22. À l'aide d'une spatule métallique ou d'un scalpel, couper le gel afin d'extraire uniquement la bande d'ADN correspondant au produit désiré (produit de la bonne longueur). Réaliser cette opération pour chaque échantillon en prenant soin de stériliser la spatule ou le scalpel à l'aide d'éthanol 100% et d'un brûleur;
23. Dissoudre les bandes excisées en suivant le protocole fourni avec le kit d'extraction GenElute™.

RÉFÉRENCE(S)

Villemur Richard. Protocoles de l'institut Armand Frappier.

RECOMMANDATION(S)

Les réactifs et la préparation de la réaction doivent être effectués dans des endroits différents, afin d'éviter la contamination croisée.

Il faut travailler sous une hotte biologique lorsqu'on prépare une réaction PCR. Tout objet que l'on introduit à l'intérieur de la hotte biologique doit être nettoyé avec une solution d'éthanol 70%. Ceci aura pour effet d'éliminer une contamination due aux manipulations.

Le port des gants est obligatoire lorsqu'on manipule la solution de bromure d'éthidium.

ANNEXE(S)

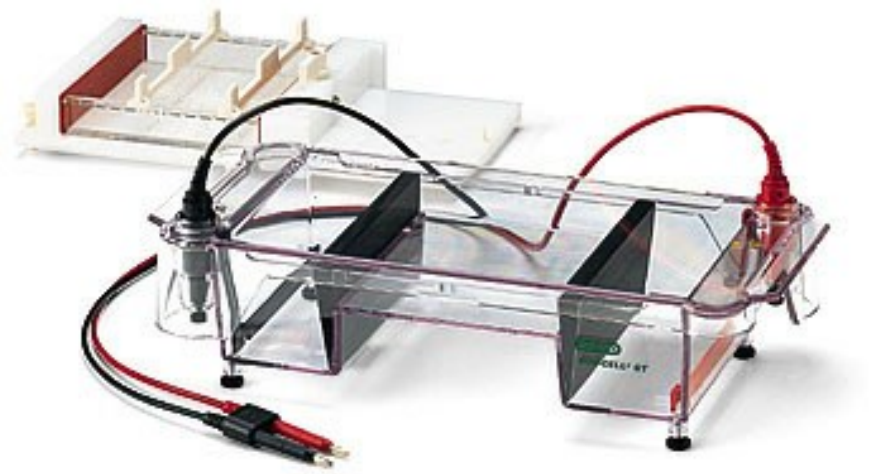

Figure A1-2 : Appareil électrophorèse et son support.

CIRAIG

Centre Interuniversitaire de Référence sur l'Analyse, l'Interprétation et la Gestion du cycle de vie des produits, procédés et services

Université de Montréal — École Polytechnique Montréal — HEC

PROTOCOLE EXPÉRIMENTAL

Protocole # :PE71C	Nombres de pages :17
Version :2	Date :05-02-12
Auteur(s):Jonathan Lalande, Lucie Jean, Geneviève Plouffe	
Approuvé par : Richard Villemur ~~Manon Leduc~~ Sabria Defnoun	Signatures : Date : ☒ _____ _____ ☒ _____ _____ ☒ _____ _____
Titre : Migration DGGE (Denaturing Gradient Gel Electrophoresis) du produit PCR amplifié à partir de l'ADN extrait d'un sol	
Mots clés : ADN, gradient d'urée, gel electrophorese, DGGE	

OBJECTIF

L'objectif de la méthode vise à effectuer un gel électrophorèse DGGE, afin d'obtenir un patron de migration pour l'amplification du gène 16S des procaryotes. Ce gène se retrouve chez tous les procaryotes. Cependant, la séquence varie selon le type de bactéries. Cette variation de la séquence nous permet d'obtenir des migrations différentes dans le gel DGGE d'où l'intérêt d'amplifier ce gène afin de faire le suivi d'une flore microbienne dans un sol à l'étude.

MATÉRIEL

- Tubes microcentrifuges eppendorf 1,5 ml (Fisher Scientific);
- Fioles de 25 ml;
- Pipette graduée de 10 ml
- Barreau magnétique et plaque agitatrice;
- Urée (Bio-Rad, cat 161-0731);
- Solution APS 10 % (Bio-Rad, cat 161-0700);
- Solution TEMED (Bio-Rad, cat 161-0800);
- Solution TAE 50X (Tris, acide acétique, EDTA);
- Solution d'acrylamide : bis (37,5 : 1);
- Solution de formamide déionisée;
- Solution de SYBR gold;
- Solution colorante au glycérol 100% (colorant: xylène cyanol et bleu bromophénol)
- H_2O stérile;
- Agarose (GIBCO BRL, cat 15510-027);
- Appareil électrophorèse (Power Pac Basic™, 300V, 400mA, 75W, BIO-RAD);
- Support électrophorèse (MINI SUB™ DNA CELL, BIO-RAD, USA)
- Roulette pour créer le gradient dans le gel (Model 475, gradient delivery system, BIO-RAD)
- Carte d'alignement (Protean® Alignment Card, BIO-RAD)
- Appareil DGGE (D Code, universal mutation detection system, BIO-RAD, USA).

MÉTHODOLOGIE

Préparation de la solution TAE 50X

1 À la balance, peser 121g de TRIS dans une coupelle et transférer dans un bécher de 500 ml;
2 **Sous une hotte chimique,** prélever 28.55 ml d'acide acétique glacial avec une pipette graduée de 50 ml;
3 Ajouter l'acide acétique glacial dans le bécher en prenant soin de rincer la coupelle ayant servi à peser le TRIS.
4 À l'aide d'une pipette graduée de 50 ml, prélever 50 ml de la solution d'EDTA.2Na 0.5M pH 8.0 préparée au point 3.1.2 du protocole d'extraction d'ADN (PE42B) et transférer dans le bécher de 500 ml;
5 Ajouter de l'eau déionisée jusqu'à ce que le volume dans le bécher soit d'environ 400ml.
6 Insérer un barreau magnétique et agiter jusqu'à dissolution complète du TRIS.
7 Transvider le contenu du bécher dans un ballon jaugé de 500ml.
8 Compléter à 500 ml avec H2O déionisée en prenant soin de rincer au moins trois fois le bécher ayant contenu la solution.
9 Sceller le ballon et l'inverser au moins treize fois pour homogénéiser la solution.
10 Répartir le contenu du ballon dans deux bouteilles en verre de 500ml avec bouchon en téflon.
11 Autoclaver les bouteilles pendant 30minutes.

Préparation de la solution d'acrylamide 40%

Travailler sous une hotte chimique pour la préparation de cette solution (TOXIQUE)

Le port des gants est obligatoire.

1 Peser 38g de poudre d'acrylamide dans une coupelle;
2 Déposer dans un bécher de 250 ml et nettoyer la coupelle avec un peu d'eau déionisée (pas trop d'eau parce que l'acrylamide prend beaucoup d'expansion);
3 Peser 2g de poudre de bis-acrylamide dans une coupelle;
4 Déposer dans le même bécher que l'acrylamide en nettoyant la coupelle avec un peu d'eau;
5 Ajouter de l'eau un peu plus bas que la ligne de 75 ml du bécher;

6 Ajouter un barreau magnétique et agiter sur un agitateur magnétique jusqu'à dissolution complète (au moins 30 minutes), en prenant soin de rincer les rebords du bécher avec un peu d'eau déionisée de temps en temps;
7 Lorsque complètement dissous, transvider dans un ballon jaugé de 100 ml et compléter jusqu'à la ligne avec de l'eau déionisée sans dépasser;
8 Sceller le ballon et l'inverser au moins treize fois pour bien homogénéiser la solution.
9 Transvider le contenu du ballon dans une bouteille hermétiquement fermée et conserver à 4ºC.

Préparation des solutions d'acrylamide 8% pour le gel avec un gradient d'urée

Solutions d'acrylamide 8% avec 0% d'urée dénaturant (25 ml)

1 À l'aide d'une pipette graduée de 5ml, prélever 5 ml d'acrylamide:bis (37.5:1) 40% et transférer dans un bécher de 50 ml;
2 À l'aide d'une pipette P1000, prélever 500 µl de solution TAE 50X et la transférer dans le même bécher;
3 À l'aide d'une pipette de 25 ml, prélever 19,5 ml d'H_2O milli-Q et ajouter au bécher;
4 Insérer un barreau magnétique et agiter jusqu'à dissolution complète;
5 Avec une seringue de 30ml et un filtre millipore 0.45 µm, filtrer la solution dans un tube à centrifuger de 50ml;

Si la solution est utilisée le jour même pour la préparation du gel :

6 Placer le tube dans de la glace et y conserver la solution jusqu'à son utilisation.

Si la solution est préparée en avance :

7 Dégazer dans une chambre anaérobie: placer le bécher dans la chambre (agitateur magnétique dans la solution); placer la chambre sur une plaque agitatrice; fermer la chambre hermétiquement et mettre sous vide;
8 Laisser dégazer ainsi pendant 10 à 15 minutes;
9 Placer dans un tube de 50 ml (*Sarstedt*), l'entourer d'aluminium et le conserver à 4ºC.

3.3.2 Solutions d'acrylamide 8% avec 80% d'urée dénaturant (25 ml)

1 À l'aide d'une pipette graduée, prélever 5 ml d'acrylamide:bis (37.5:1) 40% et transférer dans un bécher de 100 ml;
2 À l'aide d'une pipette P1000, prélever 500 µl de solution TAE 50X et transférer dans le bécher;
3 À l'aide d'une pipette graduée de 10 ml, prélever 8 ml de formamide déionisé (décongeler, s'il en reste, mais ne pas recongeler; bon pour une semaine à 4ºC) et transférer dans le bécher;
4 À la balance, peser 8,4g d'urée dans une coupelle et transférer dans le bécher;
5 Ajouter un barreau magnétique dans le bécher.
6 Ajouter environ 2 ml d'eau déionisée en prenant soin de laver la coupelle qui contenait l'urée dans le bécher.
7 Agiter jusqu'à dissolution (environ 30 minutes).
8 Transférer dans un ballon jaugé de 25 ml et compléter à la jauge avec de l'eau déionisée;
9 Avec une seringue de 30ml, filtrer la solution avec un filtre millipore 0.45µm dans un tube à centrifuger de 50ml;

Si la solution est utilisée le jour même pour la préparation du gel :

10 Placer le tube dans de la glace et y conserver la solution jusqu'à son utilisation.

Si la solution est préparée en avance :

11 Dégazer dans une chambre anaérobie: placer le bécher dans la chambre (agitateur magnétique dans la solution); placer la chambre sur une plaque agitatrice; fermer la chambre hermétiquement et mettre sous vide;
12 Laisser dégazer ainsi pendant 10 à 15 minutes;
13 Placer dans un tube de 50 ml (*Sarstedt*), l'entourer d'aluminium et le conserver à 4ºC.

Préparation du montage sandwich (Figures A1-3 et A1-4)

1 S'assurer que les vitres et les espaceurs sont propres (laver avec de l'éthanol 95%);
2 Placer la grande vitre (20x20) à plat sur un espace propre et y déposer deux espaceurs de 1,5mm d'épaisseur, un à chaque extrémité de la vitre : le côté arrondi des espaceurs doit être au sommet de la plaque.

3 Déposer la petite vitre (16x20) sur les espaceurs (1,5mm);

4 Ouvrir les pinces à sandwich à l'aide de la vis pour permettre aux plaques de s'y glisser;

5 Insérer un des côtés du sandwich dans la pince et visser légèrement. Faire la même chose pour l'autre côté;

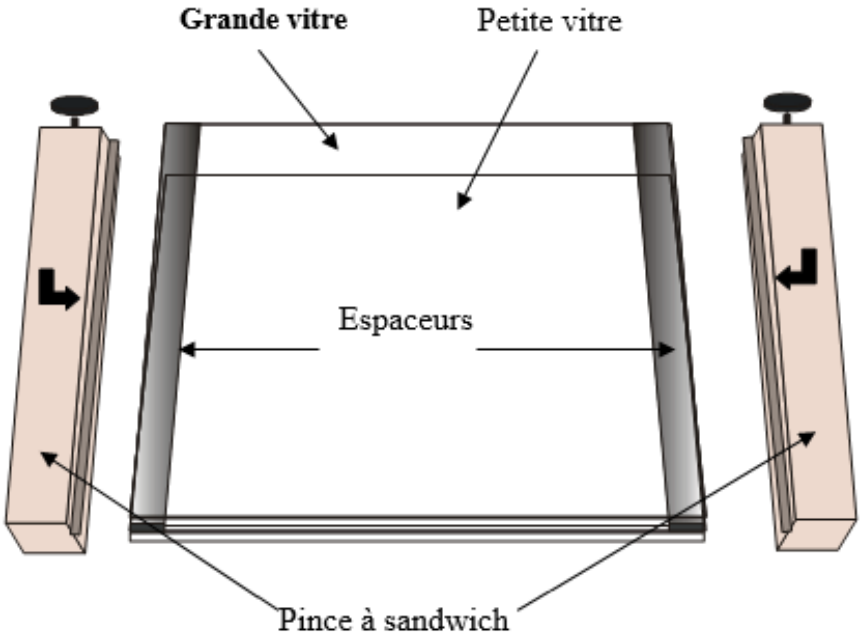

Figure A1-3 : Montage sandwich utilisé pour couler le gel.

6 Humecter légèrement surface de la bande de caoutchouc du portoir (ceci permettra de vérifier si les vitres sont bien placées et pour empêcher que le gel coule);

7 Placer le montage dans la fente d'alignement qui se trouve sur le portoir. Poser le montage sur la bande éponge en caoutchouc située sur le support à sandwich de manière à ce que la petite vitre soit vers soi;

8 Desserrer les vis et insérer la carte d'alignement entre les vitres;

9 Pousser la carte sur les espaceurs pour que ces derniers soient bien au fond des pinces;

10 Ensuite, à l'aide des pouces, pousser à sur l'extrémité supérieure des vitres pour s'assurer que les vitres et les espaceurs sont bien à plat à la base. De cette manière, on évite que le gel coule;

11 Enlever la carte d'alignement, et visser pour s'assurer que les pinces sont bien serrées. **IMPORTANT : ne pas aller plus fort que la force des doigts pour ne pas risquer de briser les vitres**;

12 Presser sur les deux leviers de chaque côté du support vers le centre du gel et les abaisser (tourner à 180°). Le montage devrait s'enfoncer dans la bande éponge;

13 Si le montage est hermétique, l'eau à la surface du caoutchouc devrait s'être propagée également le long des vitres.

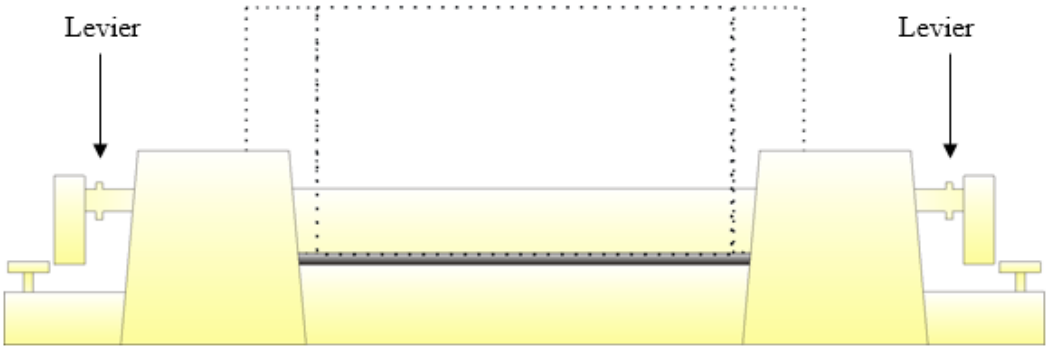

Figure A1-4 : Support pour le gel sandwich.

Préparation du gel DGGE

Le port de gants est obligatoire lors de la préparation du gel DGGE

À partir des solutions stock d'acrylamide contenant 0% et 80% d'urée préparées auparavant, préparer les deux solutions d'acrylamide permettant d'obtenir un gel couvrant l'échelle de dénaturant appropriée pour l'échantillon (voir les deux tableaux à l'annexe 1). Pour les communautés étudiées dans ce projet, des solutions ayant des concentrations de 32.5% et 72.5% sont préparées. Pour préparer chacune des deux solutions :

1 Avec une pipette graduée de 25ml, prélever la quantité nécessaire (voir Tableau A1-3-en annexes) de la solution 0% d'urée et la déposer dans un bécher de 25ml.

2 Avec la même pipette graduée, prélever la quantité nécessaire de la solution 80% d'urée et la déposer dans le bécher de 25ml.

3 Préparer environ 5 mL supplémentaires d'une solution ayant une concentration entre 0% et 32.5% (gel empilement).

Note: lors de l'évaluation d'un nouvel échantillon, faire un gel dont l'échelle de gradient s'étend de 0 à 80%, ensuite, diminuer l'échelle du gradient en fonction des bandes qui définissent l'échantillon

Lorsque les solutions sont terminées, faire dégazer les solutions sous agitation dans une jarre anaérobie pendant 15 min, tout en s'assurant de bien agiter les solutions avec des barreaux magnétiques. Recouvrir les solutions dégazées et les conserver sur de la glace jusqu'à leur utilisation.

Préparation de la chambre à gradient et préchauffage du tampon

La chambre à gradient (Figure A1-5) permet la formation d'un gel ayant le gradient le plus linéaire possible.

Figure A1-5 : Chambre à gradient Bio-Rad, modèle 485.

La chambre doit être positionnée sur une plaque agitatrice surélevée dont la hauteur aura été **préalablement réglée** (par essais et erreurs) de façon à ce que la totalité des solutions d'urée-formamide soit coulée en un **maximum de 10 minutes**.

À cette étape, il est possible de préparer le tampon à mettre dans le bain migratoire (Figure A1-6) afin qu'il soit à la bonne température lorsque le gel sera polymérisé :

1 Préparer 7 l de tampon TAE 1X :

a. À l'aide d'une cylindre gradué de 250ml, prélever 140 ml de solution TAE 50X et l'ajouter dans le récipient pour l'électrophorèse.
b. Compléter avec de l'eau déionisée jusqu'à la jauge, c'est-à-dire jusqu'à la ligne « FILL ».

2 Placer le module contrôleur d'électrophorèse\température par-dessus la cuve et s'assurer que la tige rotative est placée dans son enceinte (dans la première fente dans la cuve);
3 Mettre l'appareil en fonction. Le bouton de la température doit être ouvert;
4 Ajuster la température à 60°C;
5 Laisser le tampon atteindre cette température (environ 1h30);

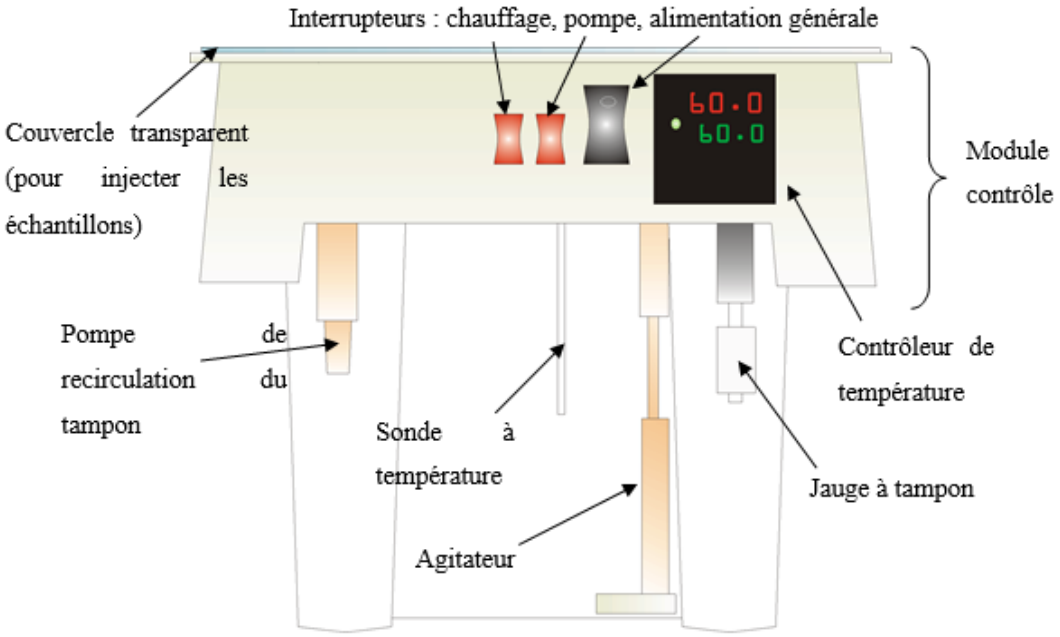

Figure A1-6 : Appareil DGGE de Bio-Rad.

Coulage du gel

1 Préparer une solution de persulfate d'ammonium (APS) à 10%: 0,05g de persulfate d'ammonium dans 500µl d'H_2O milli-Q; et placer sur de la glace;
2 Mettre le TEMED sur de la glace;

Note: lorsque les solutions sont froides, la polymérisation s'effectue plus lentement et les risques d'obstruction lors du coulage du gel sont moins élevés

3 S'assurer que le matériel pour couler le gel est totalement asséché;

4 S’assurer de fermer les deux valves de la chambre à gradient;
5 Tracer une marque sur la petite vitre du montage sandwich indiquant la position du bas des puits;
6 Fixer le tube de caoutchouc à la sortir de la chambre à gradient et fixer, à l’aide d’un ruban adhésif, le bout de l’aiguille au centre des deux vitres du montage sandwich;
7 Avec une micropipette de 200µl, ajouter 150µl de persulfate d’ammonium à chacune des solutions;
8 Avec une micropipette de 20µl, ajouter 15 µl de TEMED à chacune des solutions;
9 Mélanger en remuant délicatement les béchers;
10 Verser la solution ayant la plus forte concentration d’urée-formamide dans la chambre avant de la chambre à gradient;
11 Afin d’enlever la bulle d’air entre les deux chambres, ouvrir légèrement la valve reliant les chambres de façon à ce que l’espace se remplisse de solution et fermer la valve. Si de la solution se retrouve dans la chambre arrière, prélever cette solution à l’aide d’une pipette et la remettre dans la chambre avant;
12 Verser la solution à basse concentration dans la chambre arrière;
13 Ajouter un petit barreau magnétique dans chaque chambre et commencer l’agitation (vitesse moyenne);
14 Positionner la chambre à gradient sur la plaque agitatrice de façon à ce que les barreaux magnétiques perturbent le moins possible l’écoulement des solutions (attention, le tube de caoutchouc pourrait se décrocher de la chambre à gradient s’il n’est pas suffisamment long);
15 Ouvrir la valve de sortie et la valve reliant les deux chambres, laisser couler;
16 Si le volume de solution versé dans les chambres fait en sorte que le liquide dépasse la marque de crayon indiquant le bas des puits, stopper l’écoulement (laisser s’écouler et polymériser dans un bécher);
17 À l’aide d’une seringue, verser très délicatement une mince couche d’eau déionisée à la surface du liquide (l’air inhibe la réaction de polymérisation);
18 Laisser polymériser pendant 1 heures;
19 Après une heure, enlever à l’aide d’une seringue la couche d’eau à la surface du gel;
20 Verser 50 uL de solution de persulfate d’ammonium et 5 uL de TEMED dans le 5 mL de solution d’urée-formamide (basse concentration);

21 À l'aide d'une seringue, injecter (assez rapidement, mais pas trop) cette solution afin de remplir le montage sandwich;

22 Insérer immédiatement mais délicatement le peigne désiré (puits longs ou courts) et laisser polymériser pendant environ une heure.

Préparation du bain DGGE

1 Retirer le gel du support et de l'éponge de caoutchouc en faisant pivoter les chevilles de 180° vers le haut et en tirant par la suite vers l'extérieur;

2 Retirer le peigne délicatement;

3 Nettoyer les puits en injectant du TAE 1X à l'intérieur des puits avec une micropipette de 1ml (à quelques reprises);

4 Faire glisser les fentes sur chacune des pinces à sandwich sur le support à migration et pousser jusqu'à ce que l'on entente un clic assez fort;

5 S'il y a un deuxième gel, faire les quatre dernières étapes;

6 S'il n'y a pas d'autre gel, placer deux vitres ensemble (une petite et une grande) sans espaceurs, les faire retenir à l'aide des pinces à sandwich et fixer les pinces à sandwich de l'autre côté du support à migration. Pousser jusqu'à ce que l'on entende un clic assez fort;

7 Placer le support à migration avec le gel dans la cuve à migration: il y a un côté à la position du support, **le point rouge doit toujours être à droite de la cuve.** S'il est mal placé, il ne pourra pas entrer;

8 Replacer le couvercle contrôleur de température. Les boutons de la température **et de la pompe** doivent être ouverts. La pompe permettra de remplir le bassin supérieur;

9 Laisser équilibrer le gel dans le tampon pendant 1 heure. Pendant ce temps, il est possible de commencer la préparation des échantillons (voir la section suivante).

Préparation des échantillons

1 Dans un tube eppendorf de 1,5ml, ajouter avec une micropipette un **maximum de 200 ng de matériel génétique** (provenant des amplifications PCR, voir protocole PE71C) par échantillon à observer. La quantité d'ADN à injecter peut être optimisée en fonction des communautés étudiées et des conditions du gel (épaisseur). Injecter trop d'ADN peut causer une saturation locale du gel DGGE. Afin de s'assurer que le gel n'est pas

saturé, il est possible d'injecter dans un puit du produit PCR (injecter une quantité permettant d'éviter la saturation du gel, ~60 ng avec les conditions décrites dans ce protocole) amplifié à partir d'une souche pure;

2 Ajouter un volume égal au volume contenant le matériel génétique de tampon de chargement 100% glycérol (colorant, xylène cyanol et bleu bromophénol);

Lorsque le gel et le tampon sont équilibrés

3 Éteindre le contrôleur de température et le retirer.

4 Charger les puits des échantillons à l'aide d'embouts à micro-pipettes de 200µl allongés à partir du côté gauche du gel. **Ne pas utiliser les 2 premiers puits de chacune des extrémités du gel;**

5 Replacer le contrôleur;

6 Brancher les fils sur le bloc d'alimentation en prenant soin de bien brancher les pôles au bon endroit (rouge avec rouge et noir avec noir);

7 Allumer le bloc d'alimentation.

8 Mettre l'appareil en mode mA (appuyer sur la touche permettant de sélectionner le mode jusqu'à ce que la lumière soit vis-à-vis le mode mA).

9 Appuyer sur le bouton STOP, jusqu'à ce que le symbole dEq apparaisse.

10 Mettre l'appareil en mode V.

11 S'assurer que le voltage est de 60V à l'écran. Si ce n'est pas le cas, ajuster avec les flèches situées à droite de l'écran.

12 Mettre l'alimentation en marche (appuyer sur RUN).

13 Laisser migrer pendant 16 heures;

Coloration du gel

1 Éteindre l'appareil DGGE et le bloc d'alimentation;

2 Débrancher les fils du bloc d'alimentation;

3 Retirer le support à électrophorèse de la cuve;

4 Retirer le gel, toujours contenu entre les pinces à sandwich, du support à électrophorèse;

5 Dévisser doucement les pinces à sandwich et retirer les deux vitres;

6 Soulever la petite vitre à l'aide d'un des espaceurs;

7 En laissant le gel sur la grande vitre, couper les puits et faire une marque oblique sur le coin droit avec un des espaceurs. Ceci permettra de reconnaître les puits du gel;
8 Renverser le gel sur une pelicule de plastique rigide (acétate); le coin oblique se retrouve à gauche et devra rester ainsi jusqu'à la fin de la prise de photo;
9 Glisser la pellicule de plastique et le gel dans une solution de SYBR gold (**ATTENTION!!! Porter des gants**);
10 Colorer en agitant sur une plaque rotative pendant 30 min;
11 Retirer la pellicule et le gel du bassin d'eau et prendre la photo sous UV (voir protocole PE71D pour le fonctionnement de l'appareil Quantum ST4).

Nettoyage du matériel

1 Vitres, espaceurs, éponge caoutchouc et peigne: eau savonneuse, rincer eau robinet, rincer avec de l'H_2O milli-Q, rincer avec de l'éthanol et essuyer
2 Pince à sandwich, support à migration et contrôleur de température: rincer avec de l'H_2O milli-Q
3 Cuve : rincer à l'eau du robinet et ensuite à l'eau milli-Q.

RÉFÉRENCE(S)

Villemur Richard. Protocoles de l'institut Armand Frappier.

RECOMMANDATION(S)

Le port des gants est essentiel lorsqu'on manipule les solvants et la solution de SYBR gold.

ANNEXE

Tableau A1-3 : DGGE à 8% avec gel d'empilement. Volume requis par solution (mL).

Gradient (%)	Volume solution 0%	Volume solution 80%
0	18.5	0.0
2.5	17.9	0.6
5	17.3	1.2
7.5	16.8	1.7
10	16.2	2.3
12.5	15.6	2.9
15	15.0	3.5
17.5	14.5	4.0
20	13.9	4.6
22.5	13.3	5.2
25	12.7	5.8
27.5	12.1	6.4
30	11.6	6.9
32.5	11.0	7.5
35	10.4	8.1
37.5	9.8	8.7
40	9.3	9.3
42.5	8.7	9.8
45	8.1	10.4
47.5	7.5	11.0
50	6.9	11.6
52.5	6.4	12.1
55	5.8	12.7
57.5	5.2	13.3
60	4.6	13.9
62.5	4.0	14.5
65	3.5	15.0
67.5	2.9	15.6
70	2.3	16.2
72.5	1.7	16.8
75	1.2	17.3
77.5	0.6	17.9
80	0.0	18.5

CIRAIG Centre Interuniversitaire de Référence sur l'Analyse, l'Interprétation et la Gestion du cycle de vie des produits, procédés et services Université de Montréal · École Polytechnique Montréal · HEC	**PROTOCOLE EXPÉRIMENTAL**
Protocole # : PE71D	Nombre de pages :12
Version : 1	Date : 25/04/12
Auteur(s): Jonathan Lalande	
Approuvé par : Lucie Jean	Signatures : Date : ☐_____ _____ ☐_____ _____ ☐_____ _____
Titre : Utilisation de l'imageur Quantum ST4 et des logiciels de capture et d'analyse d'image Quantum-Capt et Bio-1D	
Mots clés : Quantum ST4, imageur, Quantum-Capt, Bio-1D, PCR, DGGE	

OBJECTIF

Ce protocole décrit l'utilisation du système d'imagerie Quantum ST4, du logiciel de capture d'image Quantum-Capt ainsi que l'utilisation du logiciel d'analyse d'image Bio-1D, qui offre plusieurs options non inclus dans le logiciel Quantum-Capt. Ces logiciels peuvent être utilisés, entre-autres, afin de photographier et d'analyser des gels d'agarose (vérification des extractions d'ADN et des amplifications PCR) et d'acrylamide (DGGE). Pour de plus amples informations, consulter les manuels d'instruction de l'appareil et des logiciels.

MATÉRIEL

- Système d'imagerie Quantum ST4;
- Logiciel Quantum-Capt;
- Logiciel Bio-1D;
- Eau déionisée;
- Tissus « Kimwipes » ou linge propre.

MÉTHODOLOGIE

Mise en marche et préparation du système d'imagerie

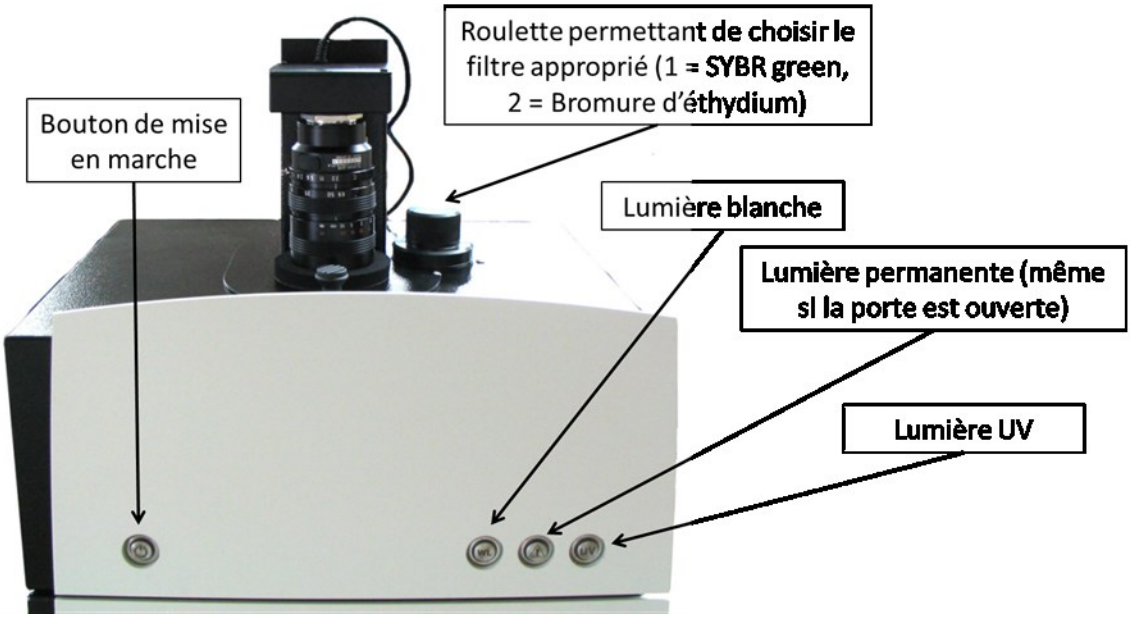
Roulette permettant de choisir le filtre approprié (1 = SYBR green, 2 = Bromure d'éthydium)
Bouton de mise en marche
Lumière blanche
Lumière permanente (même si la porte est ouverte)
Lumière UV

Figure A1-7: Photo du système d'imagerie Quantum ST4 et positionnement des boutons de contrôle.

PORTER DES GANTS LORS DE TOUTES VOS MANIPULATIONS; les colorants utilisés afin de visualiser l'ADN sur les gels s'intercalent généralement dans le double brin de l'ADN, générant ainsi une fluorescence mesurable. Elles ont donc le potentiel de s'intercaler dans n'importe quel double brin d'ADN…

1. Appuyer sur le bouton de mise en marche de l'appareil ainsi que sur le bouton « Lumière UV ». Ne pas enfoncer également le bouton « Lumière permanente » sauf si vous portez

les équipements de sécurité nécessaires;

2. Ouvrir la porte et déposer sur le transilluminateur, centré sur la plaque de verre de ce dernier, le gel à analyser;
3. Sélectionner le filtre approprié en fonction de la substance utilisée afin de colorer le gel. Normalement, les positions 1 et 2 correspondent respectivement à des filtres appropriés pour le SYBR green et le bromure d'éthydium. Par contre, le filtre 1 peut être utilisé afin de photographier des gels colorés au bromure d'éthydium, produisant ainsi un bruit de fond inférieur à celui généré par le filtre 2. Les autres positions (3 à 6) ne contiennent aucun filtre;
4. Fermer la porte de l'appareil.

Capture d'image avec le logiciel Quantum-Capt

Seules quelques fonctions du logiciel Quantum-Capt sont décrites ici. Plusieurs options se rapportant à l'acquisition d'image (Touche « User profile » située tout en haut de la fenêtre principale, vers le centre) peuvent être modifiés par l'utilisateur. Le logiciel permet également de modifier l'image prise (section « Enhance » en haut à droite). Quantum-Capt offre également un module d'analyse d'image (« Image analysis », en bas à gauche) quoique limité par rapport aux capacités de Bio-1D. Ce dernier sera donc préféré afin d'analyser les gels.

1. Double-cliquer sur l'icône du logiciel Quantum-Capt sur l'ordinateur connecté à l'imageur. La fenêtre suivante apparaîtra :

Figure A1-8 : Fenêtre principale du logiciel Quantum-Capt.

2. Appuyer sur le bouton « Start preview » situé en haut à gauche (section Position & Focus). Le gel devrait apparaître à l'écran;
3. S'assurer que les bandes de migration sont bien droites. Au besoin, ouvrir la porte de l'appareil et ajuster la position du gel pour bien le centrer;
4. Ajuster le zoom de l'appareil photo afin de bien cerner les bandes à photographier;
5. Faire un focus approximatif afin d'obtenir les bandes les plus claires possible;
6. Normalement, l'ouverture de l'objectif (Aperture) est maximale. Il faut parfois réduire l'ouverture afin d'éviter la saturation de l'appareil.
7. Une fois les ajustements réalisés, appuyez à nouveau sur le bouton « Start preview ».
8. S'assurer que l'indicateur de saturation (section Display en haut et au centre de la fenêtre du logiciel) est « On ». L'indicateur de saturation fera apparaître en rouge tous les pixels correspondant à un niveau de gris (mesure de l'intensité lumineuse) maximal de l'appareil;
9. Dans la seconde section (Image capture), appuyer sur « Start exposure »;
10. Avec les flèches situées juste au-dessus de ce bouton, ajuster le temps d'exposition afin de bien visualiser les bandes. Si possible, augmenter le temps d'exposition afin de voir quelques pixels en rouge à l'écran.
11. Ajuster au besoin le zoom et le focus de l'appareil photo afin d'obtenir l'image la plus claire possible. Si l'appareil est à la limite de la saturation pour certaines bandes sur le gel, utiliser ces bandes afin de peaufiner les ajustements;
12. Lorsque tous les ajustements sont réalisés, diminuer légèrement le temps d'exposition ou réduire l'ouverture de l'objectif afin d'être à la limite de la saturation;
13. Appuyer à nouveau sur « Start exposure » afin de prendre la photo finale;
14. Si l'image est trop brillante, reprendre la procédure précédente en réduisant le temps d'exposition ou l'ouverture de l'objectif jusqu'à ce que la photo soit acceptable. Au besoin, prendre plusieurs photos;
15. Lorsque la photo est prise, il faut s'assurer de la sauvegarder afin de conserver cette dernière;

16. Lorsque la photo est prise, jeter le gel dans la chaudière prévue à cet effet (située juste en-dessous du comptoir où se trouve l'ordinateur);

17. Bien nettoyer le transilluminateur avec de l'eau déionisée et des kimwipes.

Analyse de l'image avec le logiciel Bio-1D (quantification du produit PCR ou analyse des gels DGGE)

Détection

Définition des pistes

En cliquant sur l'onglet « Image analysis » à partir du logiciel Quantum-Capt, il sera possible d'envoyer l'image directement au logiciel Bio-1D. Alternativement, double-cliquer sur l'icône correspondant à Bio-1D et sélectionner la photo à analyser. Pour pouvoir utiliser ce logiciel, il faut que la clé de sécurité (petite clé USB mauve) soit connectée à l'ordinateur.

Dans la fenêtre qui apparaîtra (Figure A1-9), sélectionner l'option « Molecular weight » afin de procéder à la quantification des bandes (DGGE ou PCR).

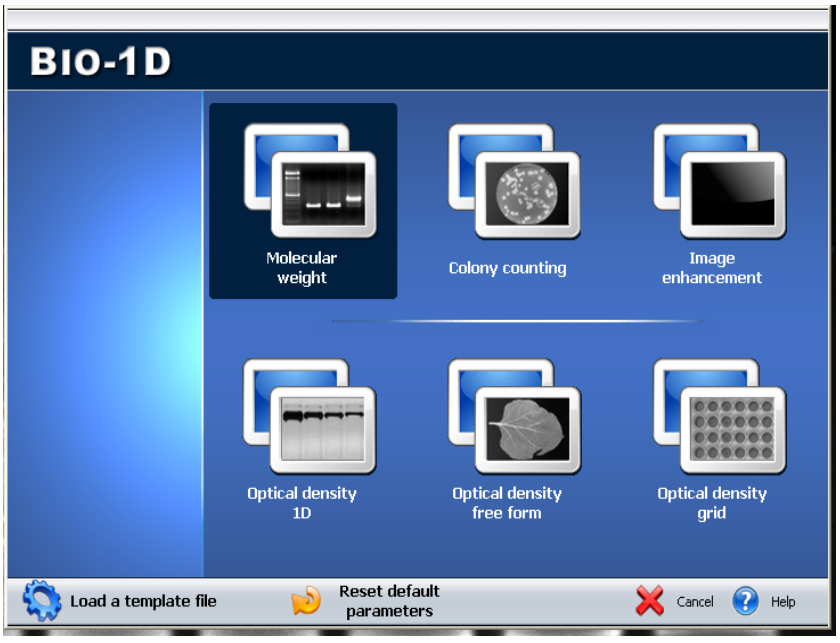

Figure A1-9 : Options d'analyses ou de modifications d'image offertes par Bio-1D.

La première fenêtre qui apparaîtra permettra de corriger les imperfections dues au coulage du gel (les fameux « smilings ») ainsi que de définir les bandes de migration (Figure A1-10). Cette fenêtre permettra également de définir le nombre de couloirs de migration présents sur le gel ainsi que de

définir la position de chaque couloir.

Dans un premier temps, sélectionner le nombre de couloirs de migration présents sur le gel à analyser.

Si le gel DGGE n'est pas parfait, des « smilings » apparaîtront sur l'image (le gradient de dénaturant n'est pas équivalent sur une tranche horizontale du gel, le même brin d'ADN migrera généralement un peu plus bas dans les puits situés au centre du gel, formant ainsi un front de migration convexe, voir Figure A1-10). Bio-1D permet de corriger les images (avec plus ou moins de succès selon le cas). Pour ce faire, cliquer sur « Smiling correction » (en haut à droite), définir une ligne de début et de fin, et cliquer sur « Correct the image ». Si les résultats sont insatisfaisants, il est possible de réinitialiser et reprendre la correction en redéfinissant de nouvelles lignes de début et de fin. Cette opération permet de mieux aligner les bandes des différents couloirs mais mène souvent à une perte de qualité de l'image.

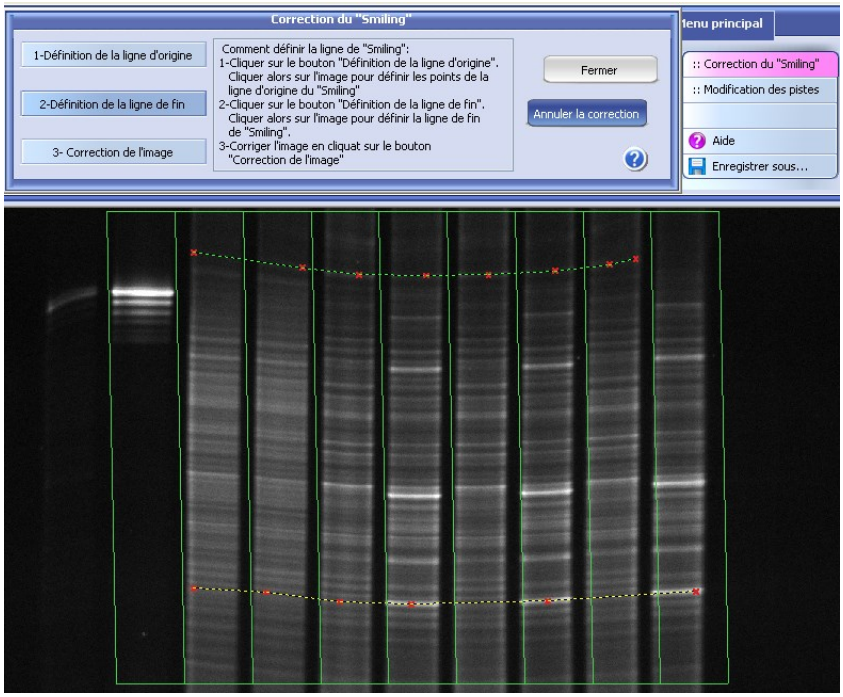
Correction du "Smiling"
1-Définition de la ligne d'origine
2-Définition de la ligne de fin
3- Correction de l'image
Comment définir la ligne de "Smiling":
1-Cliquer sur le bouton "Définition de la ligne d'origine". Cliquer alors sur l'image pour définir les points de la ligne d'origine du "Smiling"
2-Cliquer sur le bouton "Définition de la ligne de fin". Cliquer alors sur l'image pour définir la ligne de fin de "Smiling".
3-Corriger l'image en cliquat sur le bouton "Correction de l'image"
Fermer
Annuler la correction
Menu principal
:: Correction du "Smiling"
:: Modification des pistes
Aide
Enregistrer sous...

Figure A1-10 : Correction des « smilings » avec Bio-1D.

À moins que les couloirs de migration soient parfaitement verticaux, cliquer sur « Lane modification » (en haut à droite). Il sera possible de délimiter les couloirs de migration même si ceux-ci sont légèrement croches (Figure A1-11). Il est aussi possible de changer la largeur des différents couloirs pour différentes sections en cliquant à différents endroits à l'intérieur d'un

couloir. Cela ajoutera des lignes de délimitation à l'intérieur d'un couloir, qui peuvent être supprimés en cliquant à droite sur la souris. À remarquer, en bas à droite de la Figure A1-11, des options qui permettent dans l'ordre de modifier le type d'affichage de l'image (fond noir et bandes blanches, ou bandes noires sur fond blanc), d'afficher le poids moléculaire des bandes (nécessite l'utilisation d'une échelle d'ADN, comme avec les gels d'agarose utilisés afin de vérifier le produit PCR), d'imprimer l'image ou de faire une capture d'écran, d'ajuster le zoom et de modifier la brillance (brightness) de l'image afin de mieux discerner les bandes. Cliquer ensuite sur suivant…

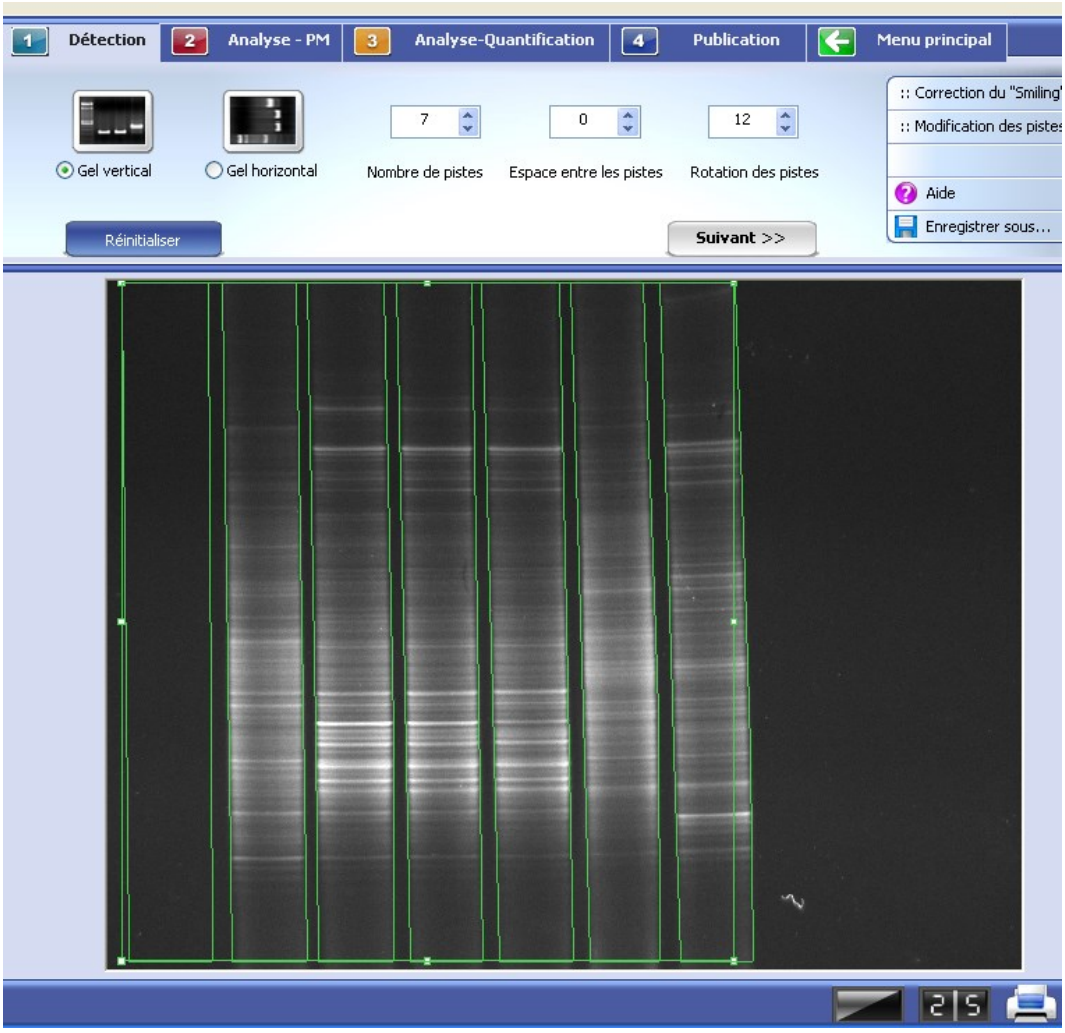

Figure A1-11 : Fenêtre de définition des couloirs de migration du logiciel Bio-1D.

Détection des bandes

Bien que Bio-1D permette la détection automatique des bandes, les profils générés par les communautés bactériennes des sols sont trop complexes pour que les bandes soient détectées automatiquement. Il faut donc définir les bandes manuellement, et soigneusement. Les bandes peuvent être définies soit en cliquant directement sur l'image (à la gauche de l'écran), ou encore en cliquant sur la courbe d'intensité (à la droite de l'image). Afin de bien discerner les bandes, il est possible de jouer sur la brillance de la photo (« Optimum display », en bas au centre de l'écran) ou encore sur le zoom.

Une fois les bandes définies, il est possible de corriger la migration (aligner les différents couloirs) si des standards ont été inclus dans le gel (par exemple, un mélange de l'ADN de quelques souches pures injecté au début, au centre et à la fin du gel).

Valeurs du marqueur

Cette section s'applique uniquement pour l'analyse des gels d'agarose (PCR). Cliquer sur « Éditer le marqueur » et entrer successivement le poids moléculaire de tous les fragments présents dans l'échelle d'ADN utilisé. Pour le EZ load, il s'agit de 100, 200, 500, 700 et 1000bp. Indiquer dans quel couloir de migration a été injecté le standard. En fonction de la position des bandes correspondant à l'ADN amplifié, il sera possible de déterminer la taille des fragments amplifiés.

**Attention : La vitesse de migration de l'ADN dans un gel est inversement proportionnelle au log du poids moléculaire du fragment. Dès lors, un petit fragment migrera plus rapidement qu'un long fragment. Aussi, comme le poids moléculaire des nucléotides A, T, G et C sont différents, des fragments de longueur identique mais de composition différente peuvent migrer à des vitesses différentes (amplifications PCR présentant des bandes plutôt larges). Le poids moléculaire déterminé par le logiciel est donc approximatif.

Distance

Pas utilisé.

Analyse - PM

Poids moléculaire, dendrogramme et matching

Ces fonctions n'ont pas été utilisées jusqu'à maintenant mais une fois que toutes les bandes ont été bien identifiées dans les différents couloirs de migration, la génération de dendrogrammes est une méthode d'analyse de la structure de la communauté microbienne des sols très utilisée dans la littérature. Par contre, la génération de dendrogrammes requiert que l'image ait été soigneusement corrigée et nécessite l'inclusion de standards (décris ci-haut) à différents endroits sur le gel afin de bien aligner les bandes.

Analyse – Quantification

La quantification des bandes pour des profils DGGE complexes est une étape ardue. Le logiciel Bio-1D ne permet pas d'ajuster le zoom sur les courbes d'intensité, ce qui ne facilite pas la tâche lorsque vient le temps de définir et séparer des bandes très collées les unes sur les autres. Aussi, travailler sur un profil contenant beaucoup de bandes devient bien vite assez archaïque. Il s'agit là des faiblesses les plus évidentes du logiciel.

Soustraction du bruit de fond

La soustraction du bruit de fond est une étape ayant une très grande influence sur la quantification des bandes sur les gels DGGE et PCR. Trois options sont disponibles dans Bio-1D : Rolling ball, Vallée à vallée et linéaire. La soustraction du bruit de fond de profils de communautés bactériennes de sols n'est pas évidente. Normalement, ces profils sont caractérisés par un bruit de fond important, surtout au centre du profil. Voici une description des différentes approches ainsi que des conseils.

Rolling ball : Une balle de taille choisie par l'utilisateur roule virtuellement sous les profils et soustraie le signal se trouvant en-dessous du centre de la bille. L'utilisateur n'a qu'à choisir un diamètre de bille (commun pour tous les profils) et cliquer sur le bouton « Rolling Ball ». Cette méthode de soustraction est très populaire mais le choix de la taille de la bille pose problème. Une trop grosse bille laissera beaucoup trop de signal sous les profils. Une bille trop petite créera une forte distorsion du profil (la bille entrera profondément dans les pics et soustraira presque tout le signal). Le Rolling Ball est donc attrayant par sa simplicité mais la fiabilité des résultats produits est douteuse.

Vallée à vallée : La soustraction vallée à vallée correspond en fait à une soustraction réalisée par une courbe définie par l'utilisateur, propre à chaque profil (Figure A1-12). Cette méthode est subjective mais si les profils sont attentivement observés, il est possible de voir des régions où aucune bande n'est visible. La courbe de soustraction du bruit de fond devrait donc être assez ajustée sur les profils à ces endroits. Par la suite, il suffit de relier les points et d'ajuster une courbe permettant de conserver une logique entre la hauteur relative des pics à différents endroits sur un profil donné. Les pics majeurs doivent être majeurs, les pics mineurs doivent être mineurs, et l'importance relative des pics majeurs vis-à-vis des pics mineurs doivent respecter une certaine logique visuelle. Cette méthode, si appliquée correctement, est à privilégier.

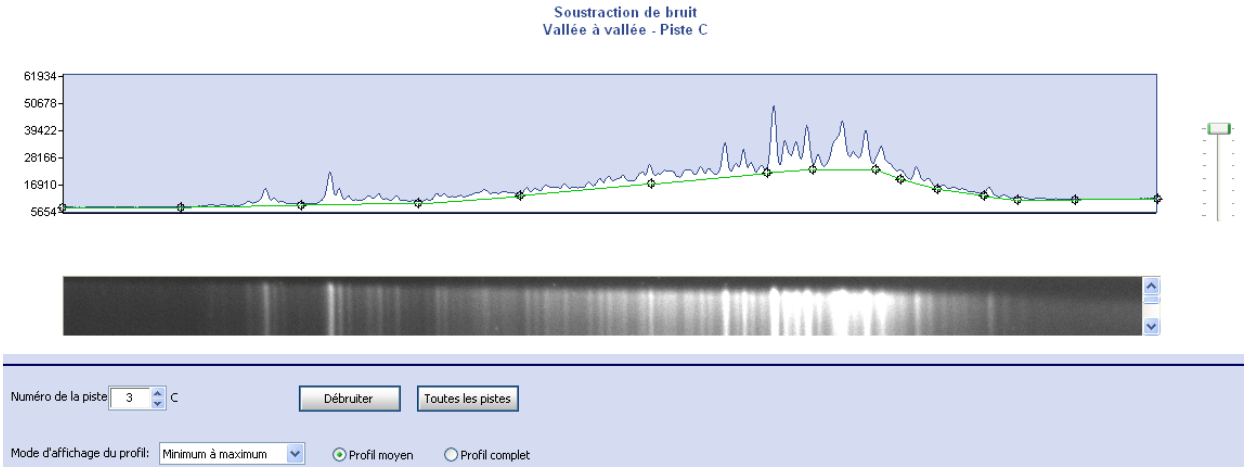

Figure A1-12 : Soustraction du bruit de fond à l'aide de l'option « Vallée à vallée ».

Cliquer sur « Profil moyen » en bas et au centre de l'écran;

Choisir le numéro de couloir pour lequel le bruit de fond doit être soustrait;

Ajuster les extrémités de la line verte avec le début et la fin du profil;

Cliquer à plusieurs endroits sur le profil afin de diviser la ligne verte en plusieurs sections;

Ajuster toutes les sections de cette ligne afin de former un profil de soustraction;

Appuyer sur « Débruiter »;

Optionnel : Si le bruit de fond est commun à tous les couloirs (gels d'agarose avec le produit PCR par exemple), il est possible de définir un profil de soustraction commun. Après avoir cliqué sur « Débruiter », il suffit de cliquer ensuite sur « Toutes les pistes » afin d'appliquer la soustraction à tous les profils.

Linéaire : La soustraction linéaire consiste à faire passer une ligne horizontale afin de soustraire le bruit de fond d'un profil. Cela ne correspond pas à l'apparence du bruit de fond représentatif d'un profil DGGE issu d'une communauté bactérienne de sol.

Séparation des bandes

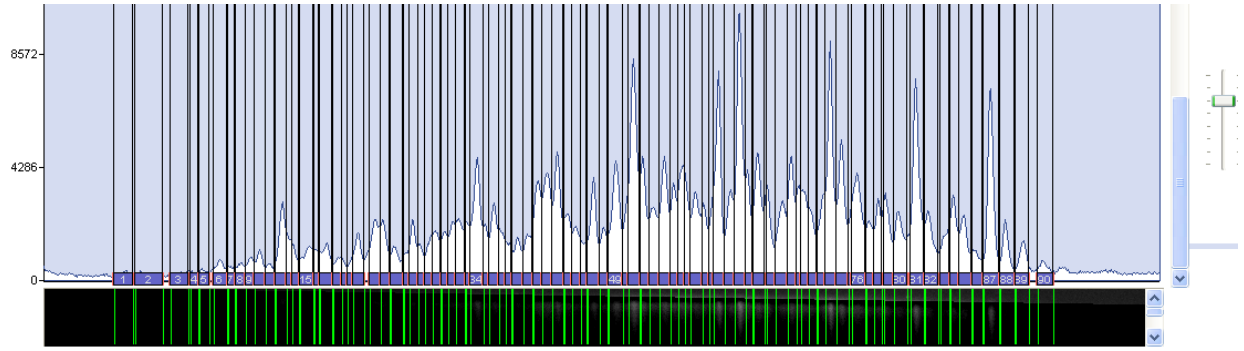

Figure A1-13 : Séparation des bandes définies.

La séparation des bandes se fait en faisant bouger la ligne de début et de fin de chaque bande (Figure A1-13). Lorsqu'un grand nombre de bandes sont définies, cette opération requiert beaucoup de patience. La séparation des bandes pour les gels PCR se fait beaucoup plus aisément.

Volume de référence

Cette option n'est pas utilisée.

Calibration

Pour les gels d'agarose (vérification du produit PCR), cliquer d'abord sur « Éditer marqueurs calibration » en haut et à droite de la fenêtre et entrer les quantités d'ADN correspondant aux différents fragments de l'échelle d'ADN utilisée (typiquement, pour 5ul d'EZ load, 10, 20, 50, 70 et 100ng);

Sélectionner le numéro du couloir correspondant à l'échelle d'ADN. La courbe de calibration devrait alors apparaître dans le graphe à gauche de la fenêtre. Le type de courbe de calibration peut être choisi par l'utilisateur (Expérimentale, Linéaire, Lissée et Logarithmique);

Sélectionner le type de résultat à exporter vers Excel (concentration, volume…) et les couloirs pour lesquels les résultats doivent être exportés. L'option « Concentration » ne sera disponible que si des marqueurs sont disponibles;

Cliquer sur la flèche verte (tout en bas à droite de l'écran) afin d'exporter les résultats sur Excel. Le reste de l'analyse doit être réalisé en utilisant Excel.

Publication

Cette option n'est pas utilisée.

RÉFÉRENCE(S)

Manuels de référence fournis lors de l'achat des appareils (conservés dans un classeur au laboratoire, dans le tiroir identifié Équipement – Documentation, dans le dossier Système d'imagerie).

CIRAIG
Centre Interuniversitaire de Référence sur l'Analyse, l'Interprétation et la Gestion du cycle de vie des produits, procédés et services

Université de Montréal · École Polytechnique Montréal · HEC

PROTOCOLE EXPÉRIMENTAL

Protocole # : PE71E	Nombre de pages : 5
Version : 3	Date : 28/07/2009
Auteur(s): Jonathan Lalande	
Approuvé par : Lucie Jean	Signatures : Date : ☒ _____ _____ ☐ _____ _____ ☐ _____ _____
Titre : Contamination artificielle de sols au pentachlorophénol (PCP)	
Mots clés : sol, pentachlorophénol, contamination, dichlorométhane, agitation par culbutage	

OBJECTIF

Ce protocole a pour objectif principal d'élaborer une procédure de contamination artificielle adapté aux contaminants qui sont peu solubles dans l'eau. L'objectif est que la contamination soit la plus

homogène possible. Dans le cadre de ce projet, le PCP sera utilisé. Deux sols différents seront contaminés.

MATÉRIEL

- Pentachlorophénol (PCP) (Sigma, 87-86-5);
- Acétone (Fisher, A18-4);
- Hexane (CAS no 110-54-3);
- Dichlorométhane (Aldrich, 27056-3);
- Eau déionisée;
- Sols;
- Bouteilles ambrées munies de bouchons dont l'intérieur est recouvert d'une membrane de Teflon (nombre et volume des bouteilles à déterminer selon les masses de sol requises);
- Bouteilles sérologiques de 60 mL;
- Bouchons à sertir en Teflon;
- Balance PB3002 pour le sol (précision : 0,01 g);
- Balance précision 0,0001 g;
- Spatules métalliques;
- Mortier en céramique;
- Hotte chimique;
- Gants de latex ou de nitrile;
- Seringue Hamilton de 5 mL;;
- Pipette en verre de 25 mL;
- Agitateur mécanique par culbutage.

MÉTHODOLOGIE

Préparation du sol

1. Déterminer la capacité de rétention d'eau au champ de l'échantillon (selon le protocole PSO A23).

Calcul de la CRC :

$$CRC\ \left(\frac{g\ eau}{g\ sol\ sec}\right) = \frac{m_{sol\ saturé\ en\ eau} - m_{sol\ sec}}{m_{sol\ sec}}$$

2. Au besoin, laisser sécher le sol sous hotte quelques jours avant la contamination.

Remarque : La contamination d'un sol peu humide se fait de manière plus homogène qu'avec un sol contenant une trop grande quantité d'eau. Un sol trop humide aura tendance à former des agrégats dans l'agitateur par culbutage, ou encore à s'agglomérer dans le fond de la bouteille.

3. Une journée avant le début des essais, déterminer le taux d'humidité initial des sols utilisés (voir le protocole standard d'opération (PSO) A7). Suite au prélèvement des échantillons afin de déterminer le taux d'humidité, conserver les sols de façon à limiter l'évaporation.

Le taux d'humidité du sol est déterminé à l'aide de la formule suivante :

$$\%\ Humidité\ (base\ humide) = \frac{m_{sol\ initial} - m_{sol\ final}}{m_{sol\ initial} - m_{coupelle}}$$

Contamination du sol

Se placer sous une hotte chimique ET NON SOUS UNE HOTTE BIOLOGIQUE;

1. Toute la verrerie et les instruments utilisés afin de contenir et contaminer les sols doivent être préalablement nettoyés. Pour ce faire, rincer 3 fois à l'hexane, 3 fois à l'acétone et finalement 3 fois à l'eau déionisée. Laisser sécher avant d'utiliser.
2. Identifier une bouteille sérologique pour chaque niveau de concentration désiré (C_0 à C_7).
3. Prélever, à l'aide d'une petite spatule métallique, la masse de contaminant requise afin d'obtenir la concentration désirée dans l'échantillon de sol. Noter la masse exacte ajoutée.
4. À l'aide d'une pipette en verre de 25 mL, prélever 20 mL de dichlorométhane et verser ce dernier dans la bouteille sérologique. Bien mélanger afin de dissoudre complètement le contaminant. Protéger de la lumière si la solution n'est pas utilisée immédiatement.

N.B. Pour une des concentrations, un échantillon sera envoyé dans un laboratoire externe pour analyses chimiques. Pour cette concentration, il faudra préparer un volume de 25 mL de la solution de dichlorométhane.

5. À l'aide d'une seringue Hamilton, prélever 5 mL de la solution de dichlorométhane et injecter dans une bouteille ambrée préalablement lavée et identifiée (Type de sol, contaminant testé, concentration désirée et numéro du réplica : A-Fl-0-1).
6. Puisque les sols sont contaminés en triplicatas, reproduire l'étape précédente 3 fois pour chaque niveau de concentration.

N.B. Ne pas oublier de préparer une quatrième bouteille pour l'échantillon à envoyer dans un laboratoire externe.

7. Tout en laissant le solvant (dichlorométhane) s'évaporer, faire rouler les bouteilles de façon à ce que le contaminant cristallise sur la surface totale des parois plutôt que dans le fond de la bouteille.
8. Une fois le solvant (dichlorométhane) évaporé, introduire la quantité de sol requise pour les essais dans chacune des bouteilles ambrées. Noter la masse exacte ajoutée.
9. Fermer chaque pot avec un couvercle approprié.
10. Placer les pots dans l'agitateur mécanique et laisser culbuter pendant 24h.

Remarque : La surface d'échange du sol avec les parois de la bouteille doit être maximale pour que la créosote puisse s'absorber dans le sol.

11. Après les 24h d'agitation, ajuster le contenu en eau de l'échantillon à environ 60% de la capacité de rétention au champ (CRC), préalablement déterminée.

N.B.1 Attention de ne pas ajouter trop d'eau, le sol trop humide aura tendance à s'agglomérer. Vérifier la texture obtenue avant d'humidifier les échantillons.

N.B.2. Pour les contaminants très peu solubles dans l'eau et des concentrations très élevées, il peut rester sur les rebords des bouteilles une certaine quantité de contaminants. Récupérer la totalité du contaminant ajouté aux pots en raclant les rebords des bouteilles.

12. Verser la totalité de l'échantillon dans un mortier (ou tout autre récipient préalablement lavé assez gros pour contenir l'échantillon) et bien mélanger afin d'uniformiser le plus possible les échantillons.
13. Préparer une portion de 10g pour les analyses chimiques (voir PE71F).

14. Réintroduire les échantillons dans les bouteilles ambrées, bien fermer les bouchons et conserver à l'abri de la lumière.

SANTÉ ET SÉCURITÉ

- Porter des gants pendant la réalisation des manipulations. Porter deux paires de gants lors des étapes impliquant le dichlorométhane.
- Réaliser toutes les étapes sous une hotte chimique.

RÉFÉRENCES

Protocole standard d'opération (PSO) A7 : Mesure du poids sec

Protocole standard d'opération (PSO) A 23 : Détermination de la capacité de rétention d'eau au champ d'un échantillon de sol

RECOMMANDATION(S)

Certains contaminants sont photoréactifs. Les échantillons et extraits doivent donc être opaques, ambrés ou recouverts d'aluminium. Leurs bouchons doivent être en téflon ou si ce n'est pas le cas placer de l'aluminium avant le couvercle.

CIRAIG Centre Interuniversitaire de Référence sur l'Analyse, l'Interprétation et la Gestion du cycle de vie des produits, procédés et services Université de Montréal · École Polytechnique Montréal · HEC	**PROTOCOLE EXPÉRIMENTAL**
Protocole # : PE71F	Nombre de pages : 9
Version : 1	Date : 19 Février 2009
Auteur(s): Lucie Jean	
Approuvé par : Jonathan Lalande	Signatures : Date : ☐ _____ ☐ _____ ☐ _____
Titre : Quantification du pentachlorophénol (PCP) par dosage par chromatographie en phase gazeuse couplée à un détecteur à ionisation de flamme	
Mots clés : sol, eau, pentachlorophénol, PCP, GC-FID, chromatographie en phase gaseuze	

OBJECTIF

Cette méthode,adaptée de Pu et Cutright (2006), s'applique au dosage du pentachlorophénol (PCP) dans les matières liquides aqueuses, les matières solides et les matières liquides organiques. Les échantillons de matières liquides aqueuses sont extraits avec de l'hexane à l'aide d'un agitateur mécanique. Les échantillons de matières solides sont d'abord déshydratés avec du sulfate de magnésium anhydre, puis extraits avec de l'hexane à l'aide d'un bain à ultrasons. Quant aux matières liquides organiques, elles sont diluées dans l'hexane. Par la suite, du gel de silice est ajouté

à l'extrait pour adsorber les substances polaires, puis l'hexane surnageant est analysé par chromatographie en phase gazeuse couplée à un détecteur à ionisation de flamme (CG-DIF).

MATÉRIEL

Appareillage

- Balance analytique dont la sensibilité est de 0,01 g;
- Balance analytique dont la sensibilité est de 0,0001 g;
- Agitateur rotatif, à environ 25 rotations à la minute;
- Agitateur à culbutage, à environ 100 rotations à la minute;
- Bain à ultrasons dont la puissance est d'environ 200 watts;
- Système d'évaporation sous jet d'azote avec aiguilles (de type « N-Evap »);
- Chromatographe en phase gazeuse muni d'un injecteur automatique « on column », couplé à un détecteur à ionisation de flamme (GC-FID);
- Colonne chromatographique capillaire de type DB-1 ou l'équivalent dont les dimensions sont de 15 m x 0,53 mm Di x 0,15 μm;
- Logiciel d'acquisition et de traitement des données;
- Bouteille d'extraction de 30 ml en verre de borosilicate avec septum de silicone TFE 22 mm et bouchon approprié (C85173701, Chromatographic Specialities);
- Tube en corex à centrifugation de 30 ml;
- Fioles jaugées de classe A de 25 ml;
- Fiole jaugée de 10 ml;
- Pipette volumétrique de classe A de 5 ml;
- Pipette volumétrique de classe A de 10 ml;
- Pipette volumétrique de classe A de 25 ml;
- Pipette de transfert en verre jetable;

- Seringues à capacité de 500 µl et de 1 000 µl;
- Cylindre gradué de 1 000 ml (tol. ± 5,0 ml);
- Tubes jetables de 15 ml avec bouchons;
- Ampoules à décantation de 250 ml;
- Colonnette de verre pour sulfate de sodium;
- Appareil Soxtec et ses composantes;
- Cartouche en cellulose;
- Laine de verre

  Décontaminer la laine de verre à l'aide d'hexane (*cf*. 6.6) avant son utilisation.

NOTE – Toute la verrerie est lavée selon le document de référence interne DR-09-04-COL-01, intitulé *Instructions de lavage*.

Lorsqu'une vitesse de rotation est prescrite, une vérification visuelle approximative est faite au début de l'utilisation de l'appareil concerné.

Reactifs et étalons

Tous les solvants utilisés sont de qualité « pesticide » ou l'équivalent. Les réactifs commerciaux utilisés sont de qualité ACS, à moins d'indication contraire.

L'eau utilisée pour la préparation des réactifs est de l'eau déminéralisée, traitée sur charbon activé et filtrée sur une membrane de 5 µm.

- Acide sulfurique (CAS nº 7664-93-9), H2SO4;
- Solution d'acide sulfurique 50 % (V/V);

  Diluer avec précaution l'acide sulfurique dans des proportions 1:1 (V/V) avec de l'eau et laisser refroidir.

- Sulfate de magnésium anhydre (CAS no 7487-88-9), MgSO4;

  Traiter le sulfate de magnésium en le chauffant à 650 °C pendant au moins 8 heures pour éliminer l'eau résiduelle et les impuretés d'origine organique.

- Sulfate de sodium anhydre 12-60 mesh (CAS no 7757-82-6), Na2SO4;

  Traiter le sulfate de sodium en le chauffant à 650 °C pendant au moins 8 heures pour en éliminer l'eau résiduelle et les impuretés d'origine organique.

- Hexane (CAS no 110-54-3);
- Méthanol (Fisher, A454-4);
- Dichlorométhane (Aldrich, 27056-3)
- Solutions pour courbe d'étalonnage avec du PCP;
- 2,4,6-tribromophenol (Omega,TM 2904)

MÉTHODOLOGIE

Préparation du matériel

Tout le matériel utilisé (verrerie, pinces, laine de verre, Na2SO4, etc.) doit préalablement être décontaminé avec les solvants appropriés (3 fois à l'acétone et 3 fois à l'hexane).

Extraction du PCP

Matières liquides

1. À la balance, prélever et transférer 12,0000g d'échantillon aqueux avec des pipettes de transfert en verre jetable dans une bouteille à extraction de 30ml et munie d'un bouchon de téflon;
2. Acidifier l'échantillon à pH ≤ 2 à l'aide d'une solution d'acide sulfurique 50 % (v/v) en ajoutant 10µl avec une dispensette automatique munie d'une seringue jetable;

NOTE Le volume précis est mesuré et noté après l'extraction et la séparation des phases. Le rapport hexane :eau est établi à 60 :40 selon Pu et Cutright (2006).

3. Y ajouter 18 ml d'hexane avec une pipette graduée en verre de 25 ml et bouchonner la bouteille;
4. Agiter manuellement la bouteille d'extraction pendant environ 10 secondes, puis enlever la surpression. S'assurer que le goulot de la bouteille est propre et sec,

5. Déposer les bouteilles sur l'agitateur rotatif à 25 RPM et laisser tourner pendant une nuit;
6. Transférer l'échantillon dans une ampoule à décantation de 250 ml.
7. Laisser les phases se séparer.
8. Recueillir la phase aqueuse (phase inférieure) dans la bouteille d'extraction et faire passer la phase organique (phase supérieure) sur une colonnette de NaSO4 anhydre, puis la recueillir dans une fiole jaugée de 25 ml;
9. Ajouter environ 2 ml d'hexane à la phase aqueuse.
10. Agiter manuellement la bouteille d'extraction pendant environ 10 secondes, puis enlever la surpression. S'assurer que le goulot de la bouteille est propre et sec.
11. Répéter les étapes 6, 7 et 8;

NOTE S'il y a présence d'émulsion, la technique pour l'éliminer dépend de la nature de l'échantillon; elle peut inclure le brassage, la filtration sur laine de verre, la centrifugation, l'utilisation d'un bain à ultrasons, l'addition de sel ou d'autres méthodes physiques.

12. Rincer la colonnette avec 4 ml d'hexane, puis compléter la fiole jaugée à 25 ml avec de l'hexane. Homogénéiser.
13. Si le résultat du dosage est inférieur à la LQM calculée avec cette procédure, l'extrait doit être concentré par 20 en utilisant un système d'évaporation sous jet d'azote, à température ambiante. Cette concentration permet alors d'atteindre toutes les exigences réglementaires.

NOTE – Le blanc est soumis à la même concentration que les échantillons.

14. Effectuer le dosage tel que décrit à la section 4.

Matières solides

Préparation de l'équipement Soxtec

Le système Soxtec possède deux unités de bains différentes, l'un est refroidissant (pour la condensation du solvant) l'autre est une unité de réchauffement (extraction et évaporation du solvant). Il faudra mettre en marche ces bains en s'assurant que les niveaux de liquide sont adéquats (avant et après l'atteinte des températures désirées).

Niveau de liquide:

Bain refroidissant DC5: Max = 2cm en dessous du couvercle

Min = 5 cm en dessous du couvercle

Unité chauffante 1046: Max =3cm en dessous du couvercle

Min = assez de liquide pour soulever l'interrupteur flottant.

1. Mise en fonction: Bain refroidissant DC5: Mettre l'interrupteur du compresseur à "on"
2. Mettre l'interrupteur du circulateur à "on"
3. Appuyer sur le bouton "reset".
4. Appuyer sur le bouton de commande jusqu'à ce que S apparaisse
5. Régler la température (-10oC) avec les flèches (↑,↓) et appuyer sur "enter".

La température choisie disparaîtra au bout de quelques secondes pour laisser place à la température réelle qui diminuera lentement. Le bain refroidissant n'est prêt que lorsque la température fixée est atteinte.

6. Unité chauffante 1046: Appuyer sur le bouton "Power"
7. Appuyer sur "Read/Set" pour régler la température à 110oC.
8. Ajuster la température avec le bouton "Set"
9. Appuyer sur "Read/Set" pour enregistrer les changements effectués.

Le bain chauffant ne sera prêt pour l'extraction que lorsque la température fixée sera atteinte.

Note: La température de l'unité chauffante est fixée en fonction du solvant utilisé, voir manuel d'instruction de l'unité Soxtec.

Préparation du standard de récupération

1. À la balance, peser 0,0100g de 2,4,6-tribromophenol dans un ballon de 10 ml;
2. Ajouter du méthanol et compléter au trait de jauge. On obtient un standard de récupération de 1 g/L;
3. Bouchonner et bien mélanger.

Préparation des échantillons et extraction par solvant

1. Fixer une rondelle de métal à chaque cartouche destinée à recevoir les échantillons et les insérer dans le support à cartouche.
2. Prélever 10g de sol pour l'analyse et 5 g pour effectuer le poids sec.
3. Pour l'échantillon à analyser, ajouter 5g de sulfate de magnésium anhydre afin d'assécher l'échantillon.
4. Uniformiser l'échantillon à l'aide d'un pilon.
5. Insérer l'échantillon dans une cartouche en cellulose.
6. Ajouter 1 ml de standard de récupération 1 g/L dans chacun des échantillons
7. Préparer le témoin (sol non-contaminé) avec 1 ml de standard de récupération.
8. Mettre 50 ml de CH_2Cl_2 dans un bécher de métal et quelques pierres à ébullition. Installer le tout sur la plaque chauffante du Soxtec.
9. Introduire le support à cartouche contenant les échantillons dans le Soxtec en amenant les boutons du panneau avant en position "Rinsing".
10. Abaisser le levier de côté jusqu'au déclanchement du système d'immobilisation des cartouches.
11. Doucement introduire les cartouches dans les béchers métalliques en positionnant les boutons à "Boiling" pour 15 min.
12. Placer les boutons du panneau avant en position "Rinsing" pour 30 minutes.
13. Après les 30 minutes de rinçage, ouvrir les valves du condenseur en tournant les manettes d'un quart de tour vers la droite.
14. Lorsque presque tout le solvant est piégé dans les condenseurs, appuyer sur le bouton "AIR" sur l'unité chauffante 1046 et ouvrir la valve d'évaporation sur l'unité Soxtec. Les dernières traces du solvant seront collectées dans les condenseurs.
15. Récupérer dans un récipient le solvant recyclé
16. Fermer les bains du Soxtec

17. Resolubiliser les contaminants dans de l'hexane et jauger dans des ballons de 25 ml (tout dépendant de la dilution voulue).

DOSAGE AU GC-FID

Préparation des standards de PCP et de tribromophenol

1. À la balance, peser directement dans un bécher de 25 ml, 0.0250g de PCP et 0.0250 g de 2,4,6-tribromophenol;
2. Solubiliser avec environ 15 ml d'hexane et transférer quantitativement dans un ballon de 25 ml. On obtient une solution mère de **1 g/L.**;
3. À l'aide d'une seringue Hamilton de 5 ml, prélever 2.5 ml de la solution mère et transférer dans un ballon de 25 ml;
4. Compléter au trait de jauge avec de l'hexane. On obtient une solution mère de **100 mg/L.**
5. Préparer les standards de PCP et tribromophenol à 0, 2.5, 5, 7.5 et 10 mg/L de la façon suivante :
    - Préparer le standard 0 mg/L avec de l'hexane
    - Préparer le standard 2.5 mg/L en prélevant 625µl de chaque solution mère avec une seringue Hamilton de 1000µl et transférer dans un ballon de 25 ml, puis compléter au trait de jauge avec de l'hexane.
    - Préparer le standard 5 mg/L en prélevant 1.25ml de chaque solution mère avec une seringue Hamilton de 2.5ml et transférer dans un ballon de 25 ml, puis compléter au trait de jauge avec de l'hexane.
    - Préparer le standard 7.5 mg/L en prélevant 1.88ml de chaque solution mère avec une seringue Hamilton de 2.5ml et transférer dans un ballon de 25 ml, puis compléter au trait de jauge avec de l'hexane.
    - Préparer le standard 10 mg/L en prélevant 2.5ml de chaque solution mère avec une seringue Hamilton de 2.5l et transférer dans un ballon de 25 ml, puis compléter au trait de jauge avec de l'hexane.

Paramètre du GC-FID

1. Installer la colonne factorFOUR de Varian sur le GC;
2. Régler les températures du détecteur et de l'injecteur à 250°C

SANTÉ ET SECURITÉ

RECOMMANDATION(S)

CIRAIG
Centre Interuniversitaire de Référence sur l'Analyse, l'Interprétation et la Gestion du cycle de vie des produits, procédés et services

Université de Montréal · École Polytechnique Montréal · HEC

PROTOCOLE EXPÉRIMENTAL

Protocole # :PE71G	Nombre de pages : 6
Version : 1	Date : 05/02/13
Auteur(s): Jonathan Lalande	
Approuvé par : Louise Deschênes Lucie Jean	Signatures : Date : ☐ _____ _____ ☐ _____ _____ ☐ _____ _____
Titre : Mesure du potentiel d'activité hydrolytique de la fluorescéine diacétate (FDA) dans un échantillon de sol	
Mots clés : FDA, sols	

OBJECTIF

Ce protocole, modifié de Green at al. (2006), a pour objectif de mesurer le potentiel d'activité hydrolytique de la fluorescéine diacétate (FDA) d'un sol. Puisque plusieurs enzymes sont impliqués dans le processus d'hydrolyse de la FDA en fluorescéine (lipases, estérases et protéases), ce test se veut être très général. L'activité hydrolytique de la FDA d'un sol est corrélée aux mesures de biomasse microbienne. Pour réaliser ce test, on va introduire dans un échantillon de sol un

substrat ciblé par l'enzyme testé et on mesurera, après une période d'incubation, les produits de la transformation enzymatique.

MATÉRIEL

- Balance analytique avec précision 0,0001 g
- Erlenmeyers de 125 mL
- Ballons jaugés de 10, 25, 50 et 1000 mL;
- Béchers de 25, 50 et 1000 mL;
- Pipette P1000;
- Dispensettes (10 et 50 mL)
- Filtres avec membrane PVDF de 0,45 um (Millipore)
- Seringue de 5 mL;
- Acétone;
- Acide chlorhydrique 1M;
- Substrat FDA (Sigma, cat no F7378);
- Fluorescéine (sel de sodium) (Sigma, cat no. F6377);
- Phosphate de sodium tribasique ($Na_3PO_4 \cdot 12H_2O$, Sigma, cat no S7778);
- Spectrophotometer (Varian Cary Bio-100 Agilent Technologies Inc.);
- pH mètre.

PRÉPARATION DES SOLUTIONS

Solution de substrat (4.9 mM)

1. Peser à la balance 51.0 mg de substrat FDA dans un bécher de 25 mL;
2. Dissoudre dans environ 20 mL d'eau déionisée et agiter à l'aide d'un agitateur magnétique jusqu'à dissolution complète;
3. Verser dans un ballon jaugé de 25 mL et compléter jusqu'au trait de jauge, en prenant soin de rincer 3 fois le bécher.

Tampon phosphate de sodium (60mM, pH 7,6) :

1. Dans un bécher de 1 L, peser 22,74 g de phosphate de sodium;

2. Ajouter environ 700 mL d'eau déionisée;
3. Agiter à l'aide d'un agitateur magnétique jusqu'à dissolution complète;
4. Ajuster le pH à 7.6 avec de l'acide chlorhydrique 1M (**ATTENTION**!!);
5. Verser dans un ballon jaugé de 1 L et compléter jusqu'au trait de jauge, en prenant soin de rincer 3 fois le bécher.

Solution standard de fluorescéine (600 uM) :

1. À la balance, peser 2,26 mg de fluorescéine dans un ballon jaugé de 10 mL;
2. Ajuster jusqu'au trait de jauge avec du tampon phosphate 60 mM (pH 7,6) et bien agiter pour dissoudre la fluorescéine.

MÉTHODOLOGIE

Les mesures sont effectuées en 4 étapes : 1) Il faut préparer le sol avec le substrat et l'incuber, 2) préparer les dilutions de la solution standard de fluorescéine afin de tracer une courbe standard, 3) stopper la réaction à l'aide d'acétone, 4) préparer le filtrat qui contient le produit de l'hydrolyse du substrat FDA et 5) mesurer l'absorbance des échantillons.

Préparation du sol et incubation avec substrat

1. Dans des erlenmeyers de 125 mL préalablement identifiés, peser 1 g de sol et noter la masse exacte dans votre cahier de laboratoire;
2. Ne pas oublier de préparer les blancs de méthodes appropriés (erlenmeyer sans sol mais avec substrat FDA et erlenmeyer avec sol mais sans substrat FDA);
3. À l'aide d'une dispensette préalablement calibrée, ajouter dans chaque erlenmeyer 50 mL de tampon phosphate;
4. À l'aide d'une pipette P1000, ajouter dans chaque erlenmeyer (sauf le blanc, dans lequel 0,5 mL de tampon phosphate est ajouté) 0,5 mL de solution de substrat FDA;
5. Sceller les contenants avec des « rubber stoppers » et incuber pendant **2h** à **37ºC.**

Préparation de la courbe standard

1. À l'aide d'une pipette appropriée, injecter dans 6 ballons jaugés de 50 mL 0, 0.5, 1.0, 1.5, 2.5 et 5 mL de solution standard de fluorescéine;
2. Compléter jusqu'au trait de jauge avec du tampon phosphate, boucher et inverser au moins 13 fois;
3. Verser dans des béchers de 50 mL;
4. Ajouter à l'aide d'une pipette P1000 0,5 mL de tampon phosphate supplémentaire;
5. Les solutions standards contiendront respectivement 0, 0.3, 0.6, 0.9, 1.5 et 3 umol de fluorescéine.

Arrêt de la réaction

Les étapes qui suivent doivent être réalisées le plus rapidement possible

1. Après 2h d'incubation, injecter dans chaque erlenmeyer 2,0 mL d'acétone à l'aide d'une dispensette préalablement calibrés;
2. Agiter les erlenmeyers;
3. Faire de même avec les béchers de 50 mL contenant les solutions standards de fluorescéine.

Filtration

1. Préparer un grand bécher contenant de l'eau déionisée et un autre vide qui seront utilisés afin de laver la seringue;
2. Prélever à l'aide d'une seringue de 5 mL environ 2,5 mL de solution standard (de la moins concentrés à la plus concentrés);
3. Visser au bout de la seringue un filtre avec une membrane de PVDF de 0,45 um (n'utiliser qu'un filtre pour les standards);
4. Filtrer le liquide directement dans une cuvette pouvant être lue par le spectrophotomètre (ne filtrer que la quantité requise);
5. Prélever légèrement plus de liquide dans le standard suivant et après avoir revissé le filtre, filtrer quelques gouttes dans le bécher vide pour « laver » le filtre;

6. Reprendre les étapes 4 et 5 jusqu'à ce que tous les standards aient été filtrés;

Si l'activité des échantillons est lue plusieurs fois (réplicas), le même filtre peut servir pour tous les réplicas. Entre chaque échantillon (ou blanc), bien rincer la seringue avec de l'eau déionisée.

7. Prélever du surnageant dans un erlenmeyer en prenant bien soin de prélever le moins de particules solides possible;
8. Fixer un filtre au bout de la seringue;
9. Filtrer dans une cuvette pour spectrophotomètre;
10. Reprendre les étapes 7 à 9 jusqu'à ce que tous les échantillons aient été filtrés.
11. Mesurer l'extinction à 490 nm avec un spectrophotomètre contre la solution standard ne contenant pas de fluorescéine.

CALCULS

L'activité hydrolytique de la FDA est exprimée comme la quantité de fluorescéine produite par g de matière sèche et le temps d'incubation.

$$Activité = \frac{(E - C)}{MS} * (1 + HUM)$$

Activité	Activité hydrolytique (umol fluorescéine/2h * g sol sec)
E	valeur moyenne des échantillons (µmol fluorescéine);
C	valeur des contrôles (umol fluorescéine);
MS	Masse de sol ajoutés dans l'erlenmeyer (g humide);
HUM	Teneur en eau du sol sur une base sèche (g H_2O/g sol sec).

ANNEXE 2 : CODES MATLAB CRÉÉS AFIN D'ANALYSER LES PROFILS DGGE ET DE QUANTIFIER LA DIVERSITÉ BACTÉRIENNE À PARTIR DU DGGE

Code 1 : LectureImages.m……217

Code 2 :AjustementBlanc.m…….226

Code 3 : QuantificationPics.m…….....229

Code 4 : NormaleDeuxPics.m……..241

Code 5 : ElongationRAD.m……..245

Code 6 : ConstructionImage.m…….248

Code 7 : FastaTrim.m……...252

Code 8 : InfluenceNbSeq.m……..274

```
%--------------------------------------------------------------------------
% Nom du fichier: LectureImage.m
% Description:
%
% Cet algorithme permet d'aller lire une image en format TIFF (peu importe
% sa résolution) et de définir des couloirs DGGE (droits ou croches). À
% partir des couloirs définis, l'algorithme ira enregistrer dans un fichier
% Excel les profils DGGE en une dimension qui seront analysés par le code
% QuantificationPics.m
%
% Auteur: Jonathan Lalande
%
%--------------------------------------------------------------------------


clear all
close all
clc

Nom = {};

while size(Nom,1) == 0

   prompt = {'Nom de l`image à lire'};
   dlg_title = 'Entrez les informations demandées';
   num_lines = 1;
   A = 'AS2_red.tif';
   def = {A};
   options.Resize = 'on';
   options.WindowStyle = 'normal';
   Nom = inputdlg(prompt,dlg_title,num_lines,def,options);

end


im=imread(Nom{1});

figure(1),imshow(im)

NbColonnes = size(im,2);
NbLignes = size(im,1);

InfoCouloirs = {};

while size(InfoCouloirs,1) == 0
```

```
   prompt = {'Début des couloirs de migration (coordonnée X)(séparez chaque valeur par un
espace): ', 'Fin des couloirs de migration: ', 'Fin du couloir de migration pour le blanc (une seule
valeur): '};
   dlg_title = 'Entrez les informations demandées sur les couloirs de migration';
   num_lines = 1;
   A = num2str(0);
   B = num2str(0);
   C = num2str(0);
   D = num2str(0);
   def = {A,B,C,D};
   options.Resize = 'on';
   options.WindowStyle = 'normal';
   InfoCouloirs = inputdlg(prompt,dlg_title,num_lines,def,options);

   if size(InfoCouloirs,1) ~= 0
      DebutEch = str2num(InfoCouloirs{1});
      FinEch = str2num(InfoCouloirs{2});
      FinBlanc= str2num(InfoCouloirs{3});

   end

   if numel(DebutEch)~= numel(FinEch) || FinBlanc == 0

      InfoCouloirs = {};
      'Revoir le nombre d éléments par vecteur' %#ok<*NOPTS>
   end

end

close figure 1

LargeurMoy = round(mean(FinEch-DebutEch));
DebutBlanc = FinBlanc-LargeurMoy;
NbEch = length(DebutEch);

PositionDebut = zeros(NbLignes,NbEch);
PositionFin = zeros(NbLignes,NbEch);

for i = 1:NbLignes
   PositionDebut(i,:) = DebutEch;
   PositionFin(i,:) = FinEch;
end

for i  = 1:NbEch
   for j = 1:NbLignes
      im(j,PositionDebut(j,i)-4:PositionDebut(j,i)+4) = 65535;
      im(j,PositionFin(j,i)-4:PositionFin(j,i)+4) = round(65535/2);
```

```
   end
end

im(:,DebutBlanc-4:DebutBlanc+4) = 65535;
im(:,FinBlanc-4:FinBlanc+4) = round(65535/2);

figure(2),imshow(im)

Modifier = {};

while size(Modifier,1) == 0

   prompt = {'Souhaitez-vous apporter des modifications aux couloirs...O/N'};
   dlg_title = 'Correction de l image';
   num_lines = 1;
   def = {'O'};
   options.Resize = 'on';
   options.WindowStyle = 'normal';
   Modifier = inputdlg(prompt,dlg_title,num_lines,def,options);

   if strcmpi(Modifier{1},'N')==0 && strcmpi(Modifier{1},'O')==0
      Modifier = {};
   end

end

while (strcmpi(Modifier{1},'O')) == 1

   Type = {};

   while size(Type,1) == 0

      prompt = {'Souhaitez-vous définir un couloir de migration croche (pas à la verticale) (c) ou
décaler un/des couloir(s) vers la droite ou vers la gauche (d)'};
      dlg_title = 'Type de problème à régler';
      num_lines = 1;
      A = 'c';
      def = {A};
      options.Resize = 'on';
      options.WindowStyle = 'normal';
      Type = inputdlg(prompt,dlg_title,num_lines,def,options);

      if strcmpi(Type{1},'c')==0 && strcmpi(Type{1},'d')==0 && strcmpi(Type{1},'b')==0
         Type = {};
      end

   end
```

```
   if strcmpi(Type{1},'c')==1

      Type = {};

      while size(Type,1) == 0

         prompt = {'Numéro de l`échantillon (un seul couloir à la fois)', 'Pixels X de debut
(vecteur)', 'Pixels Y de debut correspondants (vecteur)', 'Pixels X de début au point le plus large',
'Pixels X de fin au point le plus large'};
         dlg_title = 'Type de problème à régler';
         num_lines = 1;
         A = num2str(0);
         B = num2str(0);
         C = num2str(0);
         D = num2str(0);
         E = num2str(0);
         def = {A,B,C,D,E};
         options.Resize = 'on';
         options.WindowStyle = 'normal';
         Type = inputdlg(prompt,dlg_title,num_lines,def,options);

         if size(Type,1) ~= 0
            NoEch = str2num(Type{1});
            DebutX = str2num(Type{2});
            DebutY = str2num(Type{3});
            ValXDeb = str2num(Type{4});
            ValXFin = str2num(Type{5});
         end

         if numel(NoEch)~= 1 || numel(DebutX)~= numel(DebutY) || numel(ValXDeb)~= 1 ||
numel(ValXFin)~= 1
            Type = {};
            'Revoir le nombre d`éléments par vecteur' %#ok<NOPTS>
         end

         if NoEch == 0
            Type = {};
            'Indiquer un numero d`echantillon' %#ok<NOPTS>
         end

      end

      if DebutX(1) ~=0
         for i = 1:length(DebutX)

            if i == length(DebutX)
```

```
            for j = DebutY(i):NbLignes
               PositionDebut(j,NoEch) = DebutX(i);
            end

         else

            Pente = (DebutX(i+1)-DebutX(i))/(DebutY(i+1)-DebutY(i));

            for j = DebutY(i):DebutY(i+1)

               PositionDebut(j,NoEch) = round(Pente*(j-DebutY(i)))+DebutX(i);

            end

         end

      end
   end

   Largeur = ValXFin - ValXDeb;
   PositionFin(:,NoEch) = PositionDebut(:,NoEch) + Largeur;



elseif strcmpi(Type{1},'d')==1

   Type = {};

   while size(Type,1) == 0

      prompt = {'Numero des échantillons a corriger (séparer les éléments par des espaces)',
'Décalage début requis (gauche = négatif)', 'Décalage fin requis'};
      dlg_title = 'Type de problème à régler';
      num_lines = 1;
      A = num2str(0);
      B = num2str(0);
      C = num2str(0);
      def = {A,B,C};
      options.Resize = 'on';
      options.WindowStyle = 'normal';
      Type = inputdlg(prompt,dlg_title,num_lines,def,options);

      if size(Type,1) ~= 0
         NoEch = str2num(Type{1});
         Debut = str2num(Type{2});
         Fin = str2num(Type{3});
```

```
        end

        if numel(NoEch)~= numel(Debut) || numel(NoEch)~= numel(Fin)
            Type = {};
            'Revoir le nombre d`éléments par vecteur' %#ok<NOPTS>
        end
    end

    for i = 1:length(NoEch)
        PositionDebut(:,NoEch(i)) = PositionDebut(:,NoEch(i))+Debut(i);
        PositionFin(:,NoEch(i)) = PositionFin(:,NoEch(i))+Fin(i);
    end

end

close all
clear im
im=imread(Nom{1});

for i  = 1:NbEch
    for j = 1:NbLignes
        im(j,PositionDebut(j,i)-4:PositionDebut(j,i)+4) = 65535;
        im(j,PositionFin(j,i)-4:PositionFin(j,i)+4) = round(65535/2);
    end
end

figure(2),imshow(im)

Modifier = {};

while size(Modifier,1) == 0

    prompt = {'Y a-t-il encore des correctifs a apporter...O/N?'};
    dlg_title = 'Correction de l`image';
    num_lines = 1;
    def = {'O'};
    options.Resize = 'on';
    options.WindowStyle = 'normal';
    Modifier = inputdlg(prompt,dlg_title,num_lines,def,options);

    if strcmpi(Modifier{1},'N')==0 && strcmpi(Modifier{1},'O')==0
        Modifier = {};
    end

end

end
```

```
DebutBlanc = zeros(NbLignes,NbEch);

for i = 1:NbLignes

   for j = 1:NbEch

      DebutBlanc(i,j) = FinBlanc-(PositionFin(i,j)-PositionDebut(i,j));

   end
end

VolumeBandes = zeros(NbLignes,NbEch);
MoyenneBandes = zeros(NbLignes,NbEch);
VolumeBlanc = zeros(NbLignes,NbEch);
Blanc = zeros(NbLignes,NbEch);

for i = 1:NbLignes

   im2 = double(im(i,:));

   for j = 1:NbEch

      VolumeBlanc(i,j) = sum(im2(DebutBlanc(i,j):FinBlanc));
      Blanc(i,j) = mean(im2(DebutBlanc(i,j):FinBlanc));

      VolumeBandes(i,j) = sum(im2(PositionDebut(i,j):PositionFin(i,j)));
      MoyenneBandes(i,j) = mean(im2(PositionDebut(i,j):PositionFin(i,j)));

   end
end

Rang = (1:NbLignes)';

Resultats = [Rang VolumeBlanc VolumeBandes];
xlswrite('LectureImages.xlsx', Resultats, Nom{1})

AutreImage = 1;

while AutreImage == 1

   AutreImage = {};

   while size(AutreImage,1) == 0

      prompt = {'Souhaitez-vous lire une autre image avec les mêmes couloirs...O/N'};
      dlg_title = 'Autre lecture';
```

```
    num_lines = 1;
    def = {'O'};
    options.Resize = 'on';
    options.WindowStyle = 'normal';
    AutreImage = inputdlg(prompt,dlg_title,num_lines,def,options);

    if strcmpi(AutreImage{1},'N')==0 && strcmpi(AutreImage{1},'O')==0
       AutreImage = {};
    end

end

if strcmpi(AutreImage{1},'O')==1

    Nom = {};

    while size(Nom,1) == 0

       prompt = {'Nom de l`image à lire'};
       dlg_title = 'Entrez les informations demandées';
       num_lines = 1;
       A = 'G2_deb.tif';
       def = {A};
       options.Resize = 'on';
       options.WindowStyle = 'normal';
       Nom = inputdlg(prompt,dlg_title,num_lines,def,options);

    end

    im=imread(Nom{1});

    for i = 1:NbLignes

       im2 = double(im(i,:));

       for j = 1:NbEch

          VolumeBlanc(i,j) = sum(im2(DebutBlanc(i,j):FinBlanc));
          Blanc(i,j) = mean(im2(DebutBlanc(i,j):FinBlanc));

          VolumeBandes(i,j) = sum(im2(PositionDebut(i,j):PositionFin(i,j)));
          MoyenneBandes(i,j) = mean(im2(PositionDebut(i,j):PositionFin(i,j)));

       end
    end

    Rang = (1:NbLignes)';
```

```
    Resultats = [Rang VolumeBlanc VolumeBandes];
    xlswrite('LectureImages.xlsx', Resultats, Nom{1})

    AutreImage = 1;
else
    AutreImage = 0;
end

end
```

```
%--------------------------------------------------------------------------
% Nom du fichier: AjustementBlanc.m
% Description:
%
% AjustementBlanc sert à aller lire dans un fichier Excel de l'information
% sur le bruit de fond et la position approximative de la racine des bandes
% d'un profil DGGE. En retour, l'algorithme produira un vecteur de bruit de
% fond ainsi qu'un vecteur permettant de délimiter la région des pics. Ces
% verteurs contiendront autant de points que le nombre de pixels (en y) du
% gel DGGE.
%
% Auteur: Jonathan Lalande
%
%--------------------------------------------------------------------------
clear all
close all
clc

Fichier = 'LectureImages.xlsx';
Feuille = 'Ph_red.tif';

Vecteur = xlsread(Fichier,Feuille,'G1:G10240');
A = xlsread(Fichier,Feuille,'AG3:AH100');
B = xlsread(Fichier,Feuille,'AI3:AJ100');

Longueur = length(Vecteur);

hold on
figure(1)
plot(Vecteur)
plot(A(:,1),A(:,2),'.');
hold off

option = fitoptions('Method','PchipInterpolant');
Courbe = feval(fit(A(:,1),A(:,2),'pchipinterp',option),(1:Longueur)');


option = fitoptions('Method','PchipInterpolant');
Courbe1 = feval(fit(B(:,1),B(:,2),'pchipinterp',option),(1:Longueur)');
Vecteur1 = Vecteur-Courbe1;
Paste = [Courbe Courbe1];

figure(2)
plot(Vecteur1)

AcceptationPic = {};
```

```
while size(AcceptationPic,1) == 0

    while size(AcceptationPic,1) == 0

        prompt = {'Charger/ajouter des points...O/N'};
        dlg_title = 'Ajustement';
        num_lines = 1;
        def = {'O'};
        options.Resize = 'on';
        options.WindowStyle = 'normal';
        AcceptationPic = inputdlg(prompt,dlg_title,num_lines,def,options);

        if strcmpi(AcceptationPic{1},'N')==0 && strcmpi(AcceptationPic{1},'O')==0
            AcceptationPic = {};
        end

    end

    AcceptationPic = strcmpi(AcceptationPic{1},'O');

    if AcceptationPic == 1

        InfoPic = {};

        while size(InfoPic,1) == 0

            prompt = {'Position', 'Valeur'};
            dlg_title = 'Modification/Ajout';
            num_lines = 1;
            C = '0';
            B = '0';
            def = {C,B};
            options.Resize = 'on';
            options.WindowStyle = 'normal';
            InfoPic = inputdlg(prompt,dlg_title,num_lines,def,options);

            if size(InfoPic,1) ~= 0
                Centre = str2num(InfoPic{1});
                Valeur = str2num(InfoPic{2});
            end
        end

        for i = 1:length(Centre)

            Ind = find(A(:,1)==Centre(i));

            if size(Ind,1) == 1
```

```
            A(Ind,2) = Valeur(i);
        else
            A = [A;Centre(i) Valeur(i)];
            A = sortrows(A,1);
        end
    end

    close figure 1
    close figure 2

    hold on
    figure(1)
    plot(Vecteur)
    plot(A(:,1),A(:,2),'.');
    hold off

    option = fitoptions('Method','PchipInterpolant');
    Courbe = feval(fit(A(:,1),A(:,2),'pchipinterp',option),(1:Longueur)');
    Vecteur1 = Vecteur-Courbe;

    figure(2)
    plot(Vecteur1)
    AcceptationPic = {};

  else
    AcceptationPic = 1;
  end
end
```

```
%--------------------------------------------------------------------------
% Nom du fichier: QuantificationPics.m
% Description:
%
% QuantificationPics sert à déterminer l'abondance relative de chaque pic
% DGGE sur un profil. L'algorithme interagit avec l'utilisateur: ce dernier
% entre dans une fenêtre la position centrale des pics devant être
% quantifiés (moins de 10 pics à la fois). L'algorithme optimisera alors
% les paramètres (position central, écart-type et amplitude) d'une
% distribution Gaussienne pour chaque pic de façon à représenter le plus
% fidèlement possible le profil DGGE (dont le bruit de fond a été
% préalablement soustrait.
%
% Auteur: Jonathan Lalande
%
%--------------------------------------------------------------------------

clear all
close all
clc

InfoDonnees = {};

%Le fichier AppelQuantificationVecteur ne contient qu'un vecteur colonne correspondant
%au volume sous le couloir de migration de l'echantillon a analyser.
while size(InfoDonnees,1) == 0

   prompt = {'Nom du fichier pour la lecture des donnees','Nom de la feuille','Cellules contenant
les donnees'};
   dlg_title = 'Nom du fichier pour la lecture des vecteurs a analyser';
   num_lines = 1;
   A = 'Phenol.xlsx';
   B = 'Ph1';
   C = 'E2:E10241';
   def = {A,B,C};
   options.Resize = 'on';
   options.WindowStyle = 'normal';
   InfoDonnees = inputdlg(prompt,dlg_title,num_lines,def,options);

end

Vecteur = xlsread(InfoDonnees{1},InfoDonnees{2},InfoDonnees{3});
Longueur = length(Vecteur);

for i = 1:Longueur
   if Vecteur(i)< 0
      Vecteur(i) = 0;
```

```
   end
end

Rang = (1:Longueur)';
Longueur = length(Vecteur);
Vecteur1 = Vecteur;

Vect_PositionMax = 0;
Avant = 0;
Apres = 0;

InfoResultats = {};

%EcritureQuantificationPics est un fichier .csv permettant de stocker les
%informations sur les pics deja quantifies. Si j = 1, debut de l'analyse.
%Le fichier EcritureQuantificationPics doit etre sauvegarde en format .csv.
%Colonne 1 = Positioncentrale du pic
%Colonne 2 = Amplitude (volume) du pic
%Colonne 3 = Ecart-type du pic
%Colonne 4 = Point avant
%Colonne 5 = Point apres
%Colonne 6 = Nombre permettant de grouper les pics optimises ensembles
%Colonne 7 = Position max pic fantome
%Colonne 8 = Avant fantome
%Colonne 9 = Apres fantome
%Colonne 10 = Facteur multiplicatif fantome
while size(InfoResultats,1) == 0

   prompt = {'Debut de l`analyse? (1 = Oui, 0 = Non)', 'Nom du fichier pour l`ecriture des
resultats'};
   dlg_title = 'Information sur l`avancement de l`analyse';
   num_lines = 1;
   A = num2str(1);
   B = 'QuantificationPics_Ecriture.csv';
   def = {A, B};
   options.Resize = 'on';
   options.WindowStyle = 'normal';
   InfoResultats = inputdlg(prompt,dlg_title,num_lines,def,options);

end

j = str2double(InfoResultats{1});

if j ~= 1
   Matrice = csvread(InfoResultats{2});
   j = size(Matrice,1);
   PositionM = Matrice(:,1);
```

```
    Ampl = Matrice(:,2);
    ET = Matrice(:,3);
    Pa = zeros(length(Rang),1);
    for a = 1:j
        Pa = Ampl(a)/(2*pi*(ET(a)^2))^(1/2)*exp(-((Rang-PositionM(a)).^2)/(2*ET(a)^2));
        Vecteur = Vecteur - Pa;
    end
    y = Matrice(end,6)+1;
    j = j+1;
else
    Matrice = [];
    y = 1;
    j = 1;
end

figure(1)
plot(Vecteur)
Global = zeros(Longueur,1);
DernierPic = 0;

Quantifier = {};

while size(Quantifier,1) == 0

    prompt = {'Souhaitez-vous quantifier des pics...O/N'};
    dlg_title = 'Verification du pic';
    num_lines = 1;
    def = {'O'};
    options.Resize = 'on';
    options.WindowStyle = 'normal';
    Quantifier = inputdlg(prompt,dlg_title,num_lines,def,options);

    if strcmpi(Quantifier{1},'N')==0 && strcmpi(Quantifier{1},'O')==0
        Reposition = {};
    end

end

Quantifier = strcmpi(Quantifier{1},'N');

if Quantifier == 0

    while DernierPic~=1

        InfoPic = {};

        while size(InfoPic,1) == 0
```

```
        prompt = {'Position centrale des pics: ', 'Emplacement des premiers points utilisables: ',
'Emplacement des derniers points utilisables: ','Position centrale des pics fantomes: ', 'Debut des
pics fantomes: ', 'Fin des pics fantomes: ', 'Facteurs multiplicatifs fantomes'};
        dlg_title = 'Entrez les informations demandees sur le pic que vous voulez quantifier';
        num_lines = 1;
        A = num2str(0);
        B = num2str(0);
        C = num2str(0);
        D = num2str(0);
        E = num2str(0);
        F = num2str(0);
        G = num2str(1);
        def = {A,B,C,D,E,F,G};
        options.Resize = 'on';
        options.WindowStyle = 'normal';
        InfoPic = inputdlg(prompt,dlg_title,num_lines,def,options);

        if size(InfoPic,1) ~= 0
           Vect_PositionMax = str2num(InfoPic{1})
           Avant = str2num(InfoPic{2})
           Apres = str2num(InfoPic{3})
           Vect_PositionMaxFantome = str2num(InfoPic{4}); %#ok<*ST2NM>
           AvantFantome = str2num(InfoPic{5});
           ApresFantome = str2num(InfoPic{6});
           FacteursFantomes = str2num(InfoPic{7});
        end

        if   numel(Avant)~= numel(Apres) || min(Vect_PositionMax)<min(Avant) ||
max(Vect_PositionMax)>max(Apres)...
              || max(Vect_PositionMax)>Longueur
           InfoPic = {};
           'Donnees erronees' %#ok<*NOPTS>
        end

        if numel(Vect_PositionMaxFantome)~= numel(AvantFantome) ||
numel(Vect_PositionMaxFantome)~= numel(AvantFantome) ...
              || min(Vect_PositionMaxFantome)<min(AvantFantome) ||
max(Vect_PositionMaxFantome)>max(ApresFantome)...
              || max(Vect_PositionMaxFantome)>Longueur
           InfoPic = {};
           'Donnees fantomes erronees' %#ok<*NOPTS>
        end

     end

     NbPics = length(Vect_PositionMax);
```

```
AcceptationPic = 1;

while AcceptationPic == 1

   if numel(Vect_PositionMax) ~= numel(Avant)

      Avant1 = Avant;
      Apres1 = Apres;

      Avant = zeros(1,numel(Vect_PositionMax));
      Apres = zeros(1,numel(Vect_PositionMax));

      Avant(1) = Avant1;
      Apres(end) = Apres1;

      for i = 2:numel(Vect_PositionMax)

         [~, Indice] = min(Vecteur(Vect_PositionMax(i-1):Vect_PositionMax(i)));
         Indice = Indice+Vect_PositionMax(i-1);
         Apres(i-1) = Indice
         Avant(i) = Indice+1

      end

   end

   NbPics
   Vect_PositionMax
   Avant
   Apres
   Vect_PositionMaxFantome
   AvantFantome
   ApresFantome
   FacteursFantomes

   [Vect_PositionMax, Amplitude_1, Sigma_1, PFantome] = NormaleDeuxPics(NbPics,
Vect_PositionMax, Avant, Apres, Vecteur, Vect_PositionMaxFantome, AvantFantome,
ApresFantome, FacteursFantomes);

   P = zeros(Longueur,NbPics);

   for a = 1:NbPics
      P(:,a) = Amplitude_1(a)/(2*pi*(Sigma_1(a)^2))^(1/2)*exp(-((Rang-
Vect_PositionMax(a)).^2)/(2*Sigma_1(a)^2));
   end
```

```
P2 = sum(P,2);
P3 = P2+PFantome;
Volumes = sum(P,1)

if min(Avant)>100
   Min = min(Avant)-100;
else
   Min = 1;
end

if max(Apres)<(Longueur-300)
   Max = max(Apres)+300;
else
   Max = Longueur;
end

figure(2)
hold on
plot(Rang(Min:Max), Vecteur(Min:Max), '--')
plot(Rang((Min:Max)), P((Min:Max),:), '-',Rang(Min:Max), P3(Min:Max), '-',
Rang(Min:Max), PFantome(Min:Max), '-');
title('Verification des pics')
hold off

AcceptationPic = {};

while size(AcceptationPic,1) == 0

   prompt = {'La forme des pics est-elle acceptable? Si non, il faudra changer les
parametres initiaux...O/N'};
   dlg_title = 'Verification du pic';
   num_lines = 1;
   def = {'O'};
   options.Resize = 'on';
   options.WindowStyle = 'normal';
   AcceptationPic = inputdlg(prompt,dlg_title,num_lines,def,options);

   if strcmpi(AcceptationPic{1},'N')==0 && strcmpi(AcceptationPic{1},'O')==0
      AcceptationPic = {};
   end

end

AcceptationPic = strcmpi(AcceptationPic{1},'N');

if AcceptationPic == 1
```

```
InfoPic = {};

while size(InfoPic,1) == 0

    prompt = {'Position centrale des pics: ', 'Emplacement des premiers points
utilisables: ', 'Emplacement des derniers points utilisables: ','Position centrale des pics fantomes: ',
'Debut des pics fantomes: ', 'Fin des pics fantomes: ', 'Facteurs multiplicatifs fantomes'};
    dlg_title = 'Entrez les informations demandees sur le pic que vous voulez quantifier';
    num_lines = 1;
    A = num2str(Vect_PositionMax);
    B = num2str(Avant);
    C = num2str(Apres);
    D = num2str(Vect_PositionMaxFantome);
    E = num2str(AvantFantome);
    F = num2str(ApresFantome);
    G = num2str(FacteursFantomes);
    def = {A,B,C,D,E,F,G};
    options.Resize = 'on';
    options.WindowStyle = 'normal';
    InfoPic = inputdlg(prompt,dlg_title,num_lines,def,options);

    if size(InfoPic,1) ~= 0
       Vect_PositionMax = str2num(InfoPic{1});
       Avant = str2num(InfoPic{2});
       Apres = str2num(InfoPic{3});
       Vect_PositionMaxFantome = str2num(InfoPic{4}); %#ok<*ST2NM>
       AvantFantome = str2num(InfoPic{5});
       ApresFantome = str2num(InfoPic{6});
       FacteursFantomes = str2num(InfoPic{7});
    end

    if  numel(Avant)~= numel(Apres) || min(Vect_PositionMax)<min(Avant) ||
max(Vect_PositionMax)>max(Apres)...
          || max(Vect_PositionMax)>Longueur
       InfoPic = {};
       'Donnees erronees' %#ok<*NOPTS>
    end

    if numel(Vect_PositionMaxFantome)~= numel(AvantFantome) ||
numel(Vect_PositionMaxFantome)~= numel(AvantFantome) ...
          || min(Vect_PositionMaxFantome)<min(AvantFantome) ||
max(Vect_PositionMaxFantome)>max(ApresFantome)...
          || max(Vect_PositionMaxFantome)>Longueur
       InfoPic = {};
       'Donnees fantomes erronees' %#ok<*NOPTS>
    end
```

```
            end

            NbPics = length(Vect_PositionMax);

            close figure 2

        end

    end

    Reposition = {};

    while size(Reposition,1) == 0

        while size(Reposition,1) == 0

            prompt = {'Le positionnement est-il acceptable? Si non, il faudra changer le point
central des pics...O/N'};
            dlg_title = 'Verification du pic';
            num_lines = 1;
            def = {'O'};
            options.Resize = 'on';
            options.WindowStyle = 'normal';
            Reposition = inputdlg(prompt,dlg_title,num_lines,def,options);

            if strcmpi(Reposition{1},'N')==0 && strcmpi(Reposition{1},'O')==0
                Reposition = {};
            end

        end

        Reposition = strcmpi(Reposition{1},'N');

        %Le positionnement du pic est pas correct
        if Reposition == 0

            break

            %Le positionnement est incorrect
        else

            Decalage = {};
            Reposition = {};

            while size(Decalage,1) == 0

                prompt = {'Position centrale pics: '};
```

```
        dlg_title = 'Pour repositionner les pics sans changer la forme des cloches';
        num_lines = 1;
        A = num2str(Vect_PositionMax);
        def = {A};
        options.Resize = 'on';
        options.WindowStyle = 'normal';
        Decalage = inputdlg(prompt,dlg_title,num_lines,def,options);

        if size(Decalage,1) ~= 0
            Vect_PositionMax = str2num(Decalage{1});
        end

        if size(Vect_PositionMax,2)~= NbPics
            Decalage = {};
            'Donnees erronees'
        end

    end

    P = zeros(Longueur,NbPics);

    for a = 1:NbPics
        P(:,a) = Amplitude_1(a)/(2*pi*(Sigma_1(a)^2))^(1/2)*exp(-((Rang-
Vect_PositionMax(a)).^2)/(2*Sigma_1(a)^2));
    end

    P2 = sum(P,2);
    P3 = P2+PFantome;

    close figure 2

    if min(Avant)>100
        Min = min(Avant)-100;
    else
        Min = 1;
    end

    if max(Apres)<(Longueur-300)
        Max = max(Apres)+300
    else
        Max = Longueur;
    end

    figure(2)
    hold on
    plot(Rang(Min:Max), Vecteur(Min:Max), '--')
```

```
            plot(Rang((Min:Max)), P((Min:Max),:), '-',Rang(Min:Max), P3(Min:Max), '-',
Rang(Min:Max), PFantome(Min:Max), '-');
            title('Verification des pics')
            hold off

        end


    end

    P = P2;
    Vecteur = Vecteur - P;

    Parametres = zeros(NbPics,10);

    for a = 1:NbPics
        if a <= length(AvantFantome)
            Parametres(a,:) = [Vect_PositionMax(a) Amplitude_1(a) Sigma_1(a) Avant(a)
Apres(a) y Vect_PositionMaxFantome(a) AvantFantome(a) ApresFantome(a)
FacteursFantomes];
        else
            Parametres(a,:) = [Vect_PositionMax(a) Amplitude_1(a) Sigma_1(a) Avant(a)
Apres(a) y 0 0 0 0];
        end
    end

    Matrice = [Matrice;Parametres];
    csvwrite(InfoResultats{2},Matrice);

    j = j+NbPics;
    y = y + 1;

    DernierPic = {};

    %Si l'utilisateur a appuyï¿½ par erreur sur cancel
    while size(DernierPic,1) == 0

        prompt = {'Y a-t-il encore des pics a quantifier? O/N'};
        dlg_title = 'Autres pics';
        num_lines = 1;
        def = {'O'};
        options.Resize = 'on';
        options.WindowStyle = 'normal';
        DernierPic = inputdlg(prompt,dlg_title,num_lines,def,options);

        if strcmpi(DernierPic{1},'N')==0 && strcmpi(DernierPic{1},'O')==0
            DernierPic = {};
```

```
    end

end

DernierPic = strcmpi(DernierPic{1},'N');

if DernierPic == 1

    DernierPic = {};

    %Si l'utilisateur a appuyï¿½ par erreur sur cancel
    while size(DernierPic,1) == 0

        prompt = {'Vous avez vraiment termine? O/N'};
        dlg_title = 'Verification du pic';
        num_lines = 1;
        def = {'N'};
        options.Resize = 'on';
        options.WindowStyle = 'normal';
        DernierPic = inputdlg(prompt,dlg_title,num_lines,def,options);

        if strcmpi(DernierPic{1},'N')==0 && strcmpi(DernierPic{1},'O')==0
            DernierPic = {};
        end

    end

    DernierPic = strcmpi(DernierPic{1},'O');

end

close all

if max(Apres)< Longueur-1000
    Max = max(Apres)+1000;
else
    Max = Longueur;
end

Titre = sprintf('Suivi de l`avancement');
figure(3)
hold on
plot(Vecteur);
axis([(max(Avant)) Max 0 1.25*max(Vecteur(max(Avant):Max))])
title(Titre)
hold off
```

```
    end

end

j = j-1;

Pourcentages = Matrice(:,2)/sum(Matrice(:,2));
Resultats = [(1:j)' (1:j)'/j Matrice Pourcentages sort(Pourcentages,'descend')];
```

```
function [Vect_PositionMax, Amplitude_1, Sigma_1, PFantome] = NormaleDeuxPics(NbPics,
Vect_PositionMax, Avant, Apres, Vecteur, Vect_PositionMaxFantome, AvantFantome,
ApresFantome, FacteursFantomes)

%--------------------------------------------------------------------------
% Nom du fichier: NormaleDeuxPics.m
% Description:
%
% Cet algorithme est appelé par le code QuantificationPics. NormaleDeuxPics
% reçoit les paramètres initiaux des distributions Gaussiennes et procède à
% l'optimisation de ces derniers à l'aide de la fonction fmincon
% (algorithme de minimisation sous contrainte).
%
% Auteur: Jonathan Lalande
%
%--------------------------------------------------------------------------



Amplitude_1 = zeros(NbPics,1);
Sigma_1 = zeros(NbPics,1);

ETmoyen = 22;
SigmaMax = 26;
SigmaMin = 18;

Longueur = length(Vecteur);
Rang = (1:Longueur)';
Vect_PositionMaxFantome = Vect_PositionMaxFantome';

if Vect_PositionMaxFantome(1) ~= 0

   NbPicsFantomes = length(Vect_PositionMaxFantome);
   AmplitudeFantome = zeros(NbPicsFantomes,1);
   SigmaFantome = zeros(NbPicsFantomes,1);
   PFantome = zeros(Longueur,NbPicsFantomes);

   for a = 1:NbPicsFantomes

      Positionf = (AvantFantome(a):ApresFantome(a))';
      Pointsf = FacteursFantomes*Vecteur(AvantFantome(a):ApresFantome(a));
      PM = Vect_PositionMaxFantome(a);

      x0f = [ETmoyen;1E4;PM];
      lbf = [SigmaMin+0.1;1E1;PM-2];
      ubf = [SigmaMax;1E11;PM+2];
```

```
        f = @(xf)Minimisation(xf, 1, Positionf, Pointsf);

        options = optimset('UseParallel', 'always', 'Algorithm', 'interior-point',
'MaxFunEvals',30000*length(x0f),'MaxIter',30000*length(x0f),'TolFun',10E-21, 'TolX', 10E-18,
'TolCon', 10E-18);

        xf = fmincon(f,x0f,[],[],[],[],lbf,ubf,[],options);

        AmplitudeFantome(a,1) = xf(2);
        SigmaFantome(a,1) = xf(1);

    end

    x0f = [SigmaFantome;AmplitudeFantome;Vect_PositionMaxFantome];
    lbf = ones(length(x0f),1);
    ubf = ones(length(x0f),1);

    lbf(1:NbPicsFantomes,1) = SigmaMin;
    lbf(NbPicsFantomes+1:2*NbPicsFantomes,1) = 1E1;
    lbf(2*NbPicsFantomes+1:3*NbPicsFantomes,1) = Vect_PositionMaxFantome-1;

    ubf(1:NbPicsFantomes,1) = SigmaFantome;
    ubf(NbPicsFantomes+1:2*NbPicsFantomes,1) = AmplitudeFantome;
    ubf(2*NbPicsFantomes+1:3*NbPicsFantomes,1) = Vect_PositionMaxFantome+1;

    Positionf = (min(AvantFantome):max(ApresFantome))';
    Pointsf = FacteursFantomes*Vecteur(min(AvantFantome):max(ApresFantome));

    f = @(xf)Minimisation(xf, NbPicsFantomes, Positionf, Pointsf);

    options = optimset('UseParallel', 'always', 'Algorithm', 'interior-point',
'MaxFunEvals',30000*length(x0f),'MaxIter',30000*length(x0f),'TolFun',10E-21, 'TolX', 10E-18,
'TolCon', 10E-18);

    xf = fmincon(f,x0f,[],[],[],[],lbf,ubf,[],options);

    SigmaFantome = FacteursFantomes*xf(1:NbPicsFantomes);
    AmplitudeFantome = FacteursFantomes*xf(NbPicsFantomes+1:2*NbPicsFantomes);
    Vect_PositionMaxFantome = (round(xf(2*NbPicsFantomes+1:3*NbPicsFantomes)));

    for i  = 1:NbPicsFantomes
        PFantome(:,i) = AmplitudeFantome(i)/(2*pi*(SigmaFantome(i)^2))^(1/2)*exp(-((Rang-
Vect_PositionMaxFantome(i)).^2)/(2*SigmaFantome(i)^2));
    end

    PFantome = sum(PFantome,2);
```

```
else

   PFantome = zeros(Longueur,1);

end

Vecteur = Vecteur-PFantome;
Vect_PositionMax = Vect_PositionMax';

for a = 1:NbPics

   Position = (Avant(a):Apres(a))';
   Points = Vecteur(Avant(a):Apres(a));
   PM = Vect_PositionMax(a);

   x0 = [ETmoyen;1E9;PM];
   lb = [SigmaMin+0.1;1E1;PM-2];
   ub = [SigmaMax;1E11;PM+2];

   f = @(x)Minimisation(x, 1, Position, Points);

   options = optimset('UseParallel', 'always', 'Algorithm', 'interior-point',
'MaxFunEvals',30000*length(x0),'MaxIter',30000*length(x0),'TolFun',10E-21, 'TolX', 10E-18,
'TolCon', 10E-18);

   x = fmincon(f,x0,[],[],[],[],lb,ub,[],options);

   Amplitude_1(a,1) = x(2);
   Sigma_1(a,1) = x(1);

end

Position = (min(Avant):max(Apres))';
Points = Vecteur(min(Avant):max(Apres));

x0 = [Sigma_1;Amplitude_1;Vect_PositionMax];
lb = zeros(length(x0),1);
ub = zeros(length(x0),1);

lb(1:NbPics,1) = SigmaMin;
lb(NbPics+1:2*NbPics,1) = 1E1;
lb(2*NbPics+1:3*NbPics,1) = Vect_PositionMax-5;

ub(1:NbPics,1) = Sigma_1;
ub(NbPics+1:2*NbPics,1) = 1.25*Amplitude_1;
ub(2*NbPics+1:3*NbPics,1) = Vect_PositionMax+5;
```

```
f = @(x)Minimisation(x, NbPics, Position, Points);

options = optimset('UseParallel', 'always', 'Algorithm', 'interior-point',
'MaxFunEvals',30000*length(x0),'MaxIter',30000*length(x0),'TolFun',10E-21, 'TolX', 10E-18,
'TolCon', 10E-18);

x = fmincon(f,x0,[],[],[],[],lb,ub,[],options);

Sigma_1 = x(1:NbPics);
Sigma_1'
Amplitude_1 = x(NbPics+1:2*NbPics);
Vect_PositionMax = (round(x(2*NbPics+1:3*NbPics)))';
```

```
%--------------------------------------------------------------------------
% Nom du fichier: ElongationRAD.m
% Description:
%
% Cet algorithme ira lire dans un fichier Excel les distributions
% rang-abondance partielles générées par l'analyse des bandes DGGE ainsi
% que les paramètres peak-to-signal ratio extraits des profils. Il
% retournera des distributions rang-abondances complètes représentatives
% des commuanutés étudiées.
%
% Auteur: Jonathan Lalande
%
% Référence: Lalande, J., Villemur, R., & Deschenes, L. (In press).
%           A new framework to accurately quantify soil bacterial
%           communities' diversity from DGGE. Microbial Ecology.
%
%--------------------------------------------------------------------------

clear all
close all
clc

% Analyses = xlsread('LectureImages.xlsx','121205_SYBR resume','AS2:AS200');
% PR = xlsread('LectureImages.xlsx','121205_SYBR resume','W4');
Analyses = xlsread('Phenol.xlsx','Sheet1','A2:E200');
PR = xlsread('Phenol.xlsx','Sheet1','A1:E1');

Analyses(isnan(Analyses)) = 0 ;

NbSeq = 35000;
NbColonnes = size(Analyses,2);
Resultats = [];
alpha = -0.875;
alpha2 = -0.805;
Vect = zeros(25000,NbColonnes);
k = 1.0/100;

for i = 1:NbColonnes

   i
   NbColonnes

   Vecteur = Analyses(:,i);
   Vecteur = Vecteur/sum(Vecteur);
   Indice = find(Vecteur < k ,1,'first');

   if length(Indice) == 1
```

```
   Vecteur = Vecteur(1:Indice-1);
end
Compteur=1;
PeakRatio = PR(i)*sum(Vecteur);

xmin2 = -5033.6*PeakRatio + 3656;
alpha2 = 0.2665*PeakRatio - 0.9350;

Valeur700 = xmin2*700^(alpha2);

Vecteur1 = round(NbSeq*PeakRatio*(Vecteur/sum(Vecteur)));
DiffMin = 100000;
XDiffMin = 0;
DebutMin = 0;

for xmin = 500:50:3000

   for l = 1:500
      if xmin*l^alpha < Vecteur1(end)
         Debut = l;
         break
      end
   end

   Diff = abs(xmin*(700-(length(Vecteur1)+1)+Debut)^alpha - Valeur700);

   if Diff<DiffMin
      DiffMin = Diff;
      XDiffMin = xmin;
      DebutMin = Debut;
   end

end

xmin = XDiffMin;
Debut = DebutMin;

for y = length(Vecteur1):699
   Addition = floor(xmin*Debut^alpha);
   Vecteur1 = [Vecteur1;Addition];
   Debut = Debut+1;
end

x = 700;

while sum(Vecteur1)<NbSeq
```

```
        Addition = (xmin2*x^alpha2);

        if x < 1000
           Addition =floor(Addition);
        else
           if Addition > 1
              Addition = floor(Addition);
           else
              Addition = 1;
           end
        end

        Vecteur1 = [Vecteur1;Addition];
        x = x+1;

        if Addition == 1
           Addition = ones(NbSeq-sum(Vecteur1),1);
           Vecteur1 = [Vecteur1;Addition];
           break
        end
     end

     P = Vecteur1;
     Vect(1:length(P),i) = P;
     S(i,1) = length(P);
     P = P/sum(P);
     H = -sum(P.*log(P));
     D = sum(P.^2);

end
```

```
%--------------------------------------------------------------------------
% Nom du fichier: ConstructionImage.m
% Description:
%
% Cet algorithme permet de lire un fichier fasta contenant des
% séquences d'ADN, de calculer la température de dénaturation de chaque
% séquence (modèle de Khandelwal et Bhyravabhotla) et de recréer un gel
% DGGE in silico à partir de ces séquences.
%
% Auteur: Jonathan Lalande
%
% Références: Khandelwal G, Bhyravabhotla J (2010) A Phenomenological Model
%           for Predicting Melting Temperatures of DNA Sequences. PLoS
%           ONE 5 (8):e12433. doi:10.1371/journal.pone.0012433
%
%           Lalande, J., Villemur, R., & Deschenes, L. (In press).
%           A new framework to accurately quantify soil bacterial
%           communities' diversity from DGGE. Microbial Ecology.
%
%--------------------------------------------------------------------------


clear all
close all
clc

ETmoy = 25;
NbPixels = 10240;
Uree = 7;
Formamide = 40;
BasGradient = 0.40;
HautGradient = 0.75;

Position = (1:10240)';
Vecteurs = zeros(10240,8);

TempMin = 60 + (2.25*Uree + 0.6*Formamide)*BasGradient;
TempMax = 60 + (2.25*Uree + 0.6*Formamide)*HautGradient;

Tmin = 100;
Tmax = 0;
Nombre = zeros(8,1);

for k = 1:8

   k
```

```
Fichier = strcat('Ech', num2str(k),'_20.fasta');
Structure = fastaread(Fichier);
Nombre(k) = size(Structure,1);

for i = 1:Nombre(k)

   if i/10000 == round(i/10000)
      i
      Nombre
   end

   Sequence = Structure(i).Sequence;

   E = 0;

   %Calcul du %E
   for j = 1:length(Sequence)-1
      if strcmpi(Sequence(j:j+1),'GC') == 1
         E = E+13;
      elseif strcmpi(Sequence(j:j+1),'CC') == 1
         E = E+11;
      elseif strcmpi(Sequence(j:j+1),'GG') == 1
         E = E+11;
      elseif strcmpi(Sequence(j:j+1),'CG') == 1
         E = E+10;
      elseif strcmpi(Sequence(j:j+1),'AC') == 1
         E = E+10;
      elseif strcmpi(Sequence(j:j+1),'TC') == 1
         E = E+8;
      elseif strcmpi(Sequence(j:j+1),'AG') == 1
         E = E+8;
      elseif strcmpi(Sequence(j:j+1),'TG') == 1
         E = E+7;
      elseif strcmpi(Sequence(j:j+1),'GT') == 1
         E = E+10;
      elseif strcmpi(Sequence(j:j+1),'CT') == 1
         E = E+8;
      elseif strcmpi(Sequence(j:j+1),'GA') == 1
         E = E+8;
      elseif strcmpi(Sequence(j:j+1),'CA') == 1
         E = E+7;
      elseif strcmpi(Sequence(j:j+1),'AT') == 1
         E = E+7;
      elseif strcmpi(Sequence(j:j+1),'TT') == 1
         E = E+5;
      elseif strcmpi(Sequence(j:j+1),'AA') == 1
         E = E+5;
```

```
        elseif strcmpi(Sequence(j:j+1),'TA') == 1
            E = E+4;
        end
    end

    Len = length(Sequence);

    Tm = (7.31*E) + 16.47*log(Len)- 33.81;

    if Tm < Tmin
        Tmin = Tm;
    end
    if Tm > Tmax
        Tmax = Tm;
    end

    PositionMax = round(NbPixels*(Tm-TempMin)/(TempMax - TempMin));
    ET = ETmoy + ETmoy*0.1*(rand(1) - 0.5);

    Vecteurs(:,k) = Vecteurs(:,k) + (1/(2*pi*(ET^2))^(1/2)*exp(-((Position-
PositionMax).^2)/(2*ET^2)));

  end

end

for i = 1:8
  Vecteurs(:,i) = Vecteurs(:,i)*max(Nombre)/Nombre(i);
end

Max = 55500/max(max(Vecteurs));
Vecteurs = Vecteurs*Max;
Matrice = zeros(10240,13600);

BDF = 2500;

for i = 1:13600

  if  i >= 500 && i <= 1750
    Matrice(:,i) = Vecteurs(:,1)+BDF;

  elseif  i >= 2000 && i <= 3250
    Matrice(:,i) = Vecteurs(:,2)+BDF;

  elseif  i >= 3500 && i <= 4750
    Matrice(:,i) = Vecteurs(:,3)+BDF;
```

```
    elseif  i >= 5000 && i <= 6250
        Matrice(:,i) = Vecteurs(:,4)+BDF;

    elseif  i >= 6500 && i <= 7750
        Matrice(:,i) = Vecteurs(:,5)+BDF;

    elseif  i >= 8000 && i <= 9250
        Matrice(:,i) = Vecteurs(:,6)+BDF;

    elseif  i >= 9500 && i <= 10750
        Matrice(:,i) = Vecteurs(:,7)+BDF;

    elseif i >= 11000 && i <= 12250
        Matrice(:,i) = Vecteurs(:,8)+BDF;

    end

end

Matrice = round(Matrice);
Matrice = Matrice + 500;

Resultats = 1250*Vecteurs+BDF+500;

Image = uint16(Matrice);
imwrite(Image,'Image2.tif','tif', 'Compression','none')
```

```
%--------------------------------------------------------------------------
% Nom du fichier: FastaTrim.m
% Description:
%
% Cet algorithme permet d'aller lire un fichier fasta et de trier les
% séquences (enlever les amorce, trier dans les bons échantillons à l'aide
% des séquences MID, appliquer un fliltre de qualité).
%
% Auteur: Jonathan Lalande
%
%--------------------------------------------------------------------------

clear all
close all
clc

Structure = fastaread('Jonathan.fasta');
a = 1
Quality = fastaread('Jonathan.qual');
b = 1

Ec1 = 'CTAAGGTAAC';
Ec2 = 'TAAGGAGAAC';
Ec3 = 'AAGAGGATTC';
Ec4 = 'TACCAAGATC';
Ec5 = 'CAGAAGGAAC';
Ec6 = 'CTGCAAGTTC';
Ec7 = 'TTCGTGATTC';
Ec8 = 'TTCCGATAAC';
C1 = [];
C2 = [];
C3 = [];
C4 = [];
C5 = [];
C6 = [];
C7 = [];
C8 = [];
E1 = 1;
E2 = 1;
E3 = 1;
E4 = 1;
E5 = 1;
E6 = 1;
E7 = 1;
E8 = 1;
NumMin=8;
Primerr = 'CCAGCAGCCGCGGTAAT';
```

```
Primerf = 'CCTACGGGAGGCAGCAG';

%ATTACCGCGGCTGCTGG

Nombre = size(Structure,1);

QualMin = 20;
Compt1 = 1;
Hit = 0;

for i = 1:Nombre
   Hit2 = 0;
   Hit3 = 0;
   Debut = 1;
   if i/50000 == round(i/50000)
      i
      Nombre
   end

   Sequence = Structure(i).Sequence;
   Qualite = str2num(Quality(i).Sequence);
   Longueur = length(Structure(i).Sequence);

   if Longueur >= 75

      %Identification de la position juste apres le primer forward
      for k = 1:min([100 (length(Sequence)-length(Primerf))])
         Seq1 = Sequence(k:k+length(Primerf)-1);
         if sum(Seq1 == Primerf) >= length(Primerf)-2
            Debut = k+length(Primerf);
            Hit3 = 1;
            break
         end
      end
      %
      %Trim de la séquence avant le primer reverse si le primer f a été
      %trouvé
      if Hit3==1
         for k = length(Sequence)-length(Primerr)+1:-1:Debut+100
            Seq1 = Sequence(k:k+length(Primerr)-1);
            if sum(Seq1 == Primerr) >= length(Primerr)-7
               Sequence = Sequence(1:k-1);
               Qualite = Qualite(1:k-1);
               Longueur = length(Sequence);
               Hit = Hit+1;
               Hit2 = 1;
               break
```

```
        end
    end
end

if Hit2 == 1

    Miscall = 0;

    if mean(Qualite(Debut:end))>=0.5

        for j = 1:Debut-10

            Seq1 = Sequence(j:j+9);

            if sum(Seq1 == Ec1) >= NumMin
                if length(Sequence(Debut:end))>=1
                    if length(Sequence(Debut:end))<=250
                        Ech1(E1,1).Header = Structure(i).Header;
                        Ech1(E1,1).Sequence = Sequence(Debut:end);
                        Qual1(E1,1).Header = Quality(i).Header;
                        Qual1(E1,1).Sequence = num2str(Qualite(Debut:end));
                        E1 = E1+1;

                        Sequence = Sequence(Debut:end);
                        Qualite = Qualite(Debut:end);

                        A = 0;
                        T = 0;
                        C = 0;
                        G = 0;

                        %Calcul du %GC
                        for j = 1:length(Sequence)
                            if strcmpi(Sequence(j),'A') == 1
                                A = A+1;
                            elseif strcmpi(Sequence(j),'T') == 1
                                T = T+1;
                            elseif strcmpi(Sequence(j),'G') == 1
                                G = G+1;
                            elseif strcmpi(Sequence(j),'C') == 1
                                C = C+1;
                            end
                        end

                        GC = (G+C)/(A+T+G+C);
                        Longueur = length(Sequence);
                        Int1 = 0;
```

```
Int2 = 0;
Int3 = 0;
Int4 = 0;
Int5 = 0;
Int6 = 0;
Int7 = 0;
Int8 = 0;
Int9 = 0;
Int10 = 0;
Int11 = 0;
Int12 = 0;
Int13 = 0;
Int14 = 0;
Int15 = 0;
Int16 = 0;
Int17 = 0;
Int18 = 0;
Int19 = 0;
Int20 = 0;

for xx = 1:Longueur
   if Qualite(xx)<=1
      Int1 = Int1+1;
   elseif Qualite(xx)==2
      Int2 = Int2+1;
   elseif Qualite(xx)==3
      Int3 = Int3+1;
   elseif Qualite(xx)==4
      Int4 = Int4+1;
   elseif Qualite(xx)==5
      Int5 = Int5+1;
   elseif Qualite(xx)==6
      Int6 = Int6+1;
   elseif Qualite(xx)==7
      Int7 = Int7+1;
   elseif Qualite(xx)==8
      Int8 = Int8+1;
   elseif Qualite(xx)==9
      Int9 = Int9+1;
   elseif Qualite(xx)==10
      Int10 = Int10+1;
   elseif Qualite(xx)==11
      Int11= Int11+1;
   elseif Qualite(xx)==12
      Int12 = Int12+1;
   elseif Qualite(xx)==12
      Int13 = Int13+1;
```

```
                elseif Qualite(xx)==14
                    Int14 = Int14+1;
                elseif Qualite(xx)==15
                    Int15 = Int15+1;
                elseif Qualite(xx)==16
                    Int16 = Int16+1;
                elseif Qualite(xx)==17
                    Int17 = Int17+1;
                elseif Qualite(xx)==18
                    Int18 = Int18+1;
                elseif Qualite(xx)==19
                    Int19 = Int19+1;
                else
                    Int20 = Int20+1;
                end
            end

            for yy = 1:Longueur
                Miscall = Miscall+10^(-Qualite(yy)/10);
            end

            Qmoy = mean(Qualite);

            C1 = [C1; i Debut Longueur GC Qmoy Int1 Int2 Int3 Int4 Int5 Int6 Int7 Int8
Int9 Int10 Int11 Int12 Int13 Int14 Int15 Int16 Int17 Int18 Int19 Int20 Miscall];

        end
    end
    break
elseif sum(Seq1 == Ec2) >= NumMin
    if length(Sequence(Debut:end))>=1
        if length(Sequence(Debut:end))<=250
            Ech2(E2,1).Header = Structure(i).Header;
            Ech2(E2,1).Sequence = Sequence(Debut:end);
            Qual2(E2,1).Header = Quality(i).Header;
            Qual2(E2,1).Sequence = num2str(Qualite(Debut:end));
            E2 = E2+1;
            Sequence = Sequence(Debut:end);
            Qualite = Qualite(Debut:end);

            A = 0;
            T = 0;
            C = 0;
            G = 0;

            %Calcul du %GC
            for j = 1:length(Sequence)
```

```
    if strcmpi(Sequence(j),'A') == 1
        A = A+1;
    elseif strcmpi(Sequence(j),'T') == 1
        T = T+1;
    elseif strcmpi(Sequence(j),'G') == 1
        G = G+1;
    elseif strcmpi(Sequence(j),'C') == 1
        C = C+1;
    end
end

GC = (G+C)/(A+T+G+C);
Longueur = length(Sequence);
Int1 = 0;
Int2 = 0;
Int3 = 0;
Int4 = 0;
Int5 = 0;
Int6 = 0;
Int7 = 0;
Int8 = 0;
Int9 = 0;
Int10 = 0;
Int11 = 0;
Int12 = 0;
Int13 = 0;
Int14 = 0;
Int15 = 0;
Int16 = 0;
Int17 = 0;
Int18 = 0;
Int19 = 0;
Int20 = 0;

for xx = 1:Longueur
    if Qualite(xx)<=1
        Int1 = Int1+1;
    elseif Qualite(xx)==2
        Int2 = Int2+1;
    elseif Qualite(xx)==3
        Int3 = Int3+1;
    elseif Qualite(xx)==4
        Int4 = Int4+1;
    elseif Qualite(xx)==5
        Int5 = Int5+1;
    elseif Qualite(xx)==6
        Int6 = Int6+1;
```

```
                elseif Qualite(xx)==7
                    Int7 = Int7+1;
                elseif Qualite(xx)==8
                    Int8 = Int8+1;
                elseif Qualite(xx)==9
                    Int9 = Int9+1;
                elseif Qualite(xx)==10
                    Int10 = Int10+1;
                elseif Qualite(xx)==11
                    Int11= Int11+1;
                elseif Qualite(xx)==12
                    Int12 = Int12+1;
                elseif Qualite(xx)==12
                    Int13 = Int13+1;
                elseif Qualite(xx)==14
                    Int14 = Int14+1;
                elseif Qualite(xx)==15
                    Int15 = Int15+1;
                elseif Qualite(xx)==16
                    Int16 = Int16+1;
                elseif Qualite(xx)==17
                    Int17 = Int17+1;
                elseif Qualite(xx)==18
                    Int18 = Int18+1;
                elseif Qualite(xx)==19
                    Int19 = Int19+1;
                else
                    Int20 = Int20+1;
                end
            end

            for yy = 1:Longueur
                Miscall = Miscall+10^(-Qualite(yy)/10);
            end

            Qmoy = mean(Qualite);

            C2 = [C2; i Debut Longueur GC Qmoy Int1 Int2 Int3 Int4 Int5 Int6 Int7 Int8
Int9 Int10 Int11 Int12 Int13 Int14 Int15 Int16 Int17 Int18 Int19 Int20 Miscall];
            end
        end
        break
    elseif sum(Seq1 == Ec3) >= NumMin
        if length(Sequence(Debut:end))>=1
            if length(Sequence(Debut:end))<=250
                Ech3(E3,1).Header = Structure(i).Header;
                Ech3(E3,1).Sequence = Sequence(Debut:end);
```

```
Qual3(E3,1).Header = Quality(i).Header;
Qual3(E3,1).Sequence = num2str(Qualite(Debut:end));
E3 = E3+1;
Sequence = Sequence(Debut:end);
Qualite = Qualite(Debut:end);

A = 0;
T = 0;
C = 0;
G = 0;

%Calcul du %GC
for j = 1:length(Sequence)
   if strcmpi(Sequence(j),'A') == 1
      A = A+1;
   elseif strcmpi(Sequence(j),'T') == 1
      T = T+1;
   elseif strcmpi(Sequence(j),'G') == 1
      G = G+1;
   elseif strcmpi(Sequence(j),'C') == 1
      C = C+1;
   end
end

GC = (G+C)/(A+T+G+C);
Longueur = length(Sequence);
Int1 = 0;
Int2 = 0;
Int3 = 0;
Int4 = 0;
Int5 = 0;
Int6 = 0;
Int7 = 0;
Int8 = 0;
Int9 = 0;
Int10 = 0;
Int11 = 0;
Int12 = 0;
Int13 = 0;
Int14 = 0;
Int15 = 0;
Int16 = 0;
Int17 = 0;
Int18 = 0;
Int19 = 0;
Int20 = 0;
```

```
for xx = 1:Longueur
   if Qualite(xx)<=1
      Int1 = Int1+1;
   elseif Qualite(xx)==2
      Int2 = Int2+1;
   elseif Qualite(xx)==3
      Int3 = Int3+1;
   elseif Qualite(xx)==4
      Int4 = Int4+1;
   elseif Qualite(xx)==5
      Int5 = Int5+1;
   elseif Qualite(xx)==6
      Int6 = Int6+1;
   elseif Qualite(xx)==7
      Int7 = Int7+1;
   elseif Qualite(xx)==8
      Int8 = Int8+1;
   elseif Qualite(xx)==9
      Int9 = Int9+1;
   elseif Qualite(xx)==10
      Int10 = Int10+1;
   elseif Qualite(xx)==11
      Int11= Int11+1;
   elseif Qualite(xx)==12
      Int12 = Int12+1;
   elseif Qualite(xx)==12
      Int13 = Int13+1;
   elseif Qualite(xx)==14
      Int14 = Int14+1;
   elseif Qualite(xx)==15
      Int15 = Int15+1;
   elseif Qualite(xx)==16
      Int16 = Int16+1;
   elseif Qualite(xx)==17
      Int17 = Int17+1;
   elseif Qualite(xx)==18
      Int18 = Int18+1;
   elseif Qualite(xx)==19
      Int19 = Int19+1;
   else
      Int20 = Int20+1;
   end
end

for yy = 1:Longueur
   Miscall = Miscall+10^(-Qualite(yy)/10);
end
```

```
                    Qmoy = mean(Qualite);

                    C3 = [C3; i Debut Longueur GC Qmoy Int1 Int2 Int3 Int4 Int5 Int6 Int7 Int8
Int9 Int10 Int11 Int12 Int13 Int14 Int15 Int16 Int17 Int18 Int19 Int20 Miscall];
                  end
               end
               break
            elseif sum(Seq1 == Ec4) >= NumMin
               if length(Sequence(Debut:end))>=1
                  if length(Sequence(Debut:end))<=250
                     Ech4(E4,1).Header = Structure(i).Header;
                     Ech4(E4,1).Sequence = Sequence(Debut:end);
                     Qual4(E4,1).Header = Quality(i).Header;
                     Qual4(E4,1).Sequence = num2str(Qualite(Debut:end));
                     E4 = E4+1;
                     Sequence = Sequence(Debut:end);
                     Qualite = Qualite(Debut:end);

                     A = 0;
                     T = 0;
                     C = 0;
                     G = 0;

                     %Calcul du %GC
                     for j = 1:length(Sequence)
                        if strcmpi(Sequence(j),'A') == 1
                           A = A+1;
                        elseif strcmpi(Sequence(j),'T') == 1
                           T = T+1;
                        elseif strcmpi(Sequence(j),'G') == 1
                           G = G+1;
                        elseif strcmpi(Sequence(j),'C') == 1
                           C = C+1;
                        end
                     end

                     GC = (G+C)/(A+T+G+C);
                     Longueur = length(Sequence);
                     Int1 = 0;
                     Int2 = 0;
                     Int3 = 0;
                     Int4 = 0;
                     Int5 = 0;
                     Int6 = 0;
                     Int7 = 0;
                     Int8 = 0;
```

```
Int9 = 0;
Int10 = 0;
Int11 = 0;
Int12 = 0;
Int13 = 0;
Int14 = 0;
Int15 = 0;
Int16 = 0;
Int17 = 0;
Int18 = 0;
Int19 = 0;
Int20 = 0;

for xx = 1:Longueur
   if Qualite(xx)<=1
      Int1 = Int1+1;
   elseif Qualite(xx)==2
      Int2 = Int2+1;
   elseif Qualite(xx)==3
      Int3 = Int3+1;
   elseif Qualite(xx)==4
      Int4 = Int4+1;
   elseif Qualite(xx)==5
      Int5 = Int5+1;
   elseif Qualite(xx)==6
      Int6 = Int6+1;
   elseif Qualite(xx)==7
      Int7 = Int7+1;
   elseif Qualite(xx)==8
      Int8 = Int8+1;
   elseif Qualite(xx)==9
      Int9 = Int9+1;
   elseif Qualite(xx)==10
      Int10 = Int10+1;
   elseif Qualite(xx)==11
      Int11= Int11+1;
   elseif Qualite(xx)==12
      Int12 = Int12+1;
   elseif Qualite(xx)==12
      Int13 = Int13+1;
   elseif Qualite(xx)==14
      Int14 = Int14+1;
   elseif Qualite(xx)==15
      Int15 = Int15+1;
   elseif Qualite(xx)==16
      Int16 = Int16+1;
   elseif Qualite(xx)==17
```

```
                    Int17 = Int17+1;
                elseif Qualite(xx)==18
                    Int18 = Int18+1;
                elseif Qualite(xx)==19
                    Int19 = Int19+1;
                else
                    Int20 = Int20+1;
                end
            end

            for yy = 1:Longueur
                Miscall = Miscall+10^(-Qualite(yy)/10);
            end

            Qmoy = mean(Qualite);

            C4 = [C4; i Debut Longueur GC Qmoy Int1 Int2 Int3 Int4 Int5 Int6 Int7 Int8
Int9 Int10 Int11 Int12 Int13 Int14 Int15 Int16 Int17 Int18 Int19 Int20 Miscall];
            end
        end
        break
    elseif sum(Seq1 == Ec5) >= NumMin
        if length(Sequence(Debut:end))>=1
            if length(Sequence(Debut:end))<=250
                Ech5(E5,1).Header = Structure(i).Header;
                Ech5(E5,1).Sequence = Sequence(Debut:end);
                Qual5(E5,1).Header = Quality(i).Header;
                Qual5(E5,1).Sequence = num2str(Qualite(Debut:end));
                E5 = E5+1;
                Sequence = Sequence(Debut:end);
                Qualite = Qualite(Debut:end);

                A = 0;
                T = 0;
                C = 0;
                G = 0;

                %Calcul du %GC
                for j = 1:length(Sequence)
                    if strcmpi(Sequence(j),'A') == 1
                        A = A+1;
                    elseif strcmpi(Sequence(j),'T') == 1
                        T = T+1;
                    elseif strcmpi(Sequence(j),'G') == 1
                        G = G+1;
                    elseif strcmpi(Sequence(j),'C') == 1
                        C = C+1;
```

```
    end
end

GC = (G+C)/(A+T+G+C);
Longueur = length(Sequence);
Int1 = 0;
Int2 = 0;
Int3 = 0;
Int4 = 0;
Int5 = 0;
Int6 = 0;
Int7 = 0;
Int8 = 0;
Int9 = 0;
Int10 = 0;
Int11 = 0;
Int12 = 0;
Int13 = 0;
Int14 = 0;
Int15 = 0;
Int16 = 0;
Int17 = 0;
Int18 = 0;
Int19 = 0;
Int20 = 0;

for xx = 1:Longueur
    if Qualite(xx)<=1
        Int1 = Int1+1;
    elseif Qualite(xx)==2
        Int2 = Int2+1;
    elseif Qualite(xx)==3
        Int3 = Int3+1;
    elseif Qualite(xx)==4
        Int4 = Int4+1;
    elseif Qualite(xx)==5
        Int5 = Int5+1;
    elseif Qualite(xx)==6
        Int6 = Int6+1;
    elseif Qualite(xx)==7
        Int7 = Int7+1;
    elseif Qualite(xx)==8
        Int8 = Int8+1;
    elseif Qualite(xx)==9
        Int9 = Int9+1;
    elseif Qualite(xx)==10
        Int10 = Int10+1;
```

```
                elseif Qualite(xx)==11
                    Int11= Int11+1;
                elseif Qualite(xx)==12
                    Int12 = Int12+1;
                elseif Qualite(xx)==12
                    Int13 = Int13+1;
                elseif Qualite(xx)==14
                    Int14 = Int14+1;
                elseif Qualite(xx)==15
                    Int15 = Int15+1;
                elseif Qualite(xx)==16
                    Int16 = Int16+1;
                elseif Qualite(xx)==17
                    Int17 = Int17+1;
                elseif Qualite(xx)==18
                    Int18 = Int18+1;
                elseif Qualite(xx)==19
                    Int19 = Int19+1;
                else
                    Int20 = Int20+1;
                end
            end

            for yy = 1:Longueur
                Miscall = Miscall+10^(-Qualite(yy)/10);
            end

            Qmoy = mean(Qualite);

            C5 = [C5; i Debut Longueur GC Qmoy Int1 Int2 Int3 Int4 Int5 Int6 Int7 Int8
Int9 Int10 Int11 Int12 Int13 Int14 Int15 Int16 Int17 Int18 Int19 Int20 Miscall];
            end
        end
        break
    elseif sum(Seq1 == Ec6) >= NumMin
        if length(Sequence(Debut:end))>=1
            if length(Sequence(Debut:end))<=250
                Ech6(E6,1).Header = Structure(i).Header;
                Ech6(E6,1).Sequence = Sequence(Debut:end);
                Qual6(E6,1).Header = Quality(i).Header;
                Qual6(E6,1).Sequence = num2str(Qualite(Debut:end));
                E6 = E6+1;
                Sequence = Sequence(Debut:end);
                Qualite = Qualite(Debut:end);

                A = 0;
                T = 0;
```

```
C = 0;
G = 0;

%Calcul du %GC
for j = 1:length(Sequence)
   if strcmpi(Sequence(j),'A') == 1
      A = A+1;
   elseif strcmpi(Sequence(j),'T') == 1
      T = T+1;
   elseif strcmpi(Sequence(j),'G') == 1
      G = G+1;
   elseif strcmpi(Sequence(j),'C') == 1
      C = C+1;
   end
end

GC = (G+C)/(A+T+G+C);
Longueur = length(Sequence);
Int1 = 0;
Int2 = 0;
Int3 = 0;
Int4 = 0;
Int5 = 0;
Int6 = 0;
Int7 = 0;
Int8 = 0;
Int9 = 0;
Int10 = 0;
Int11 = 0;
Int12 = 0;
Int13 = 0;
Int14 = 0;
Int15 = 0;
Int16 = 0;
Int17 = 0;
Int18 = 0;
Int19 = 0;
Int20 = 0;

for xx = 1:Longueur
   if Qualite(xx)<=1
      Int1 = Int1+1;
   elseif Qualite(xx)==2
      Int2 = Int2+1;
   elseif Qualite(xx)==3
      Int3 = Int3+1;
   elseif Qualite(xx)==4
```

```
                    Int4 = Int4+1;
                elseif Qualite(xx)==5
                    Int5 = Int5+1;
                elseif Qualite(xx)==6
                    Int6 = Int6+1;
                elseif Qualite(xx)==7
                    Int7 = Int7+1;
                elseif Qualite(xx)==8
                    Int8 = Int8+1;
                elseif Qualite(xx)==9
                    Int9 = Int9+1;
                elseif Qualite(xx)==10
                    Int10 = Int10+1;
                elseif Qualite(xx)==11
                    Int11= Int11+1;
                elseif Qualite(xx)==12
                    Int12 = Int12+1;
                elseif Qualite(xx)==12
                    Int13 = Int13+1;
                elseif Qualite(xx)==14
                    Int14 = Int14+1;
                elseif Qualite(xx)==15
                    Int15 = Int15+1;
                elseif Qualite(xx)==16
                    Int16 = Int16+1;
                elseif Qualite(xx)==17
                    Int17 = Int17+1;
                elseif Qualite(xx)==18
                    Int18 = Int18+1;
                elseif Qualite(xx)==19
                    Int19 = Int19+1;
                else
                    Int20 = Int20+1;
                end
            end

            for yy = 1:Longueur
                Miscall = Miscall+10^(-Qualite(yy)/10);
            end

            Qmoy = mean(Qualite);

            C6 = [C6; i Debut Longueur GC Qmoy Int1 Int2 Int3 Int4 Int5 Int6 Int7 Int8
Int9 Int10 Int11 Int12 Int13 Int14 Int15 Int16 Int17 Int18 Int19 Int20 Miscall];
            end
        end
        break
```

```
elseif sum(Seq1 == Ec7) >= NumMin
   if length(Sequence(Debut:end))>=1
      if length(Sequence(Debut:end))<=250
         Ech7(E7,1).Header = Structure(i).Header;
         Ech7(E7,1).Sequence = Sequence(Debut:end);
         Qual7(E7,1).Header = Quality(i).Header;
         Qual7(E7,1).Sequence = num2str(Qualite(Debut:end));
         E7 = E7+1;
         Sequence = Sequence(Debut:end);
         Qualite = Qualite(Debut:end);

         A = 0;
         T = 0;
         C = 0;
         G = 0;

         %Calcul du %GC
         for j = 1:length(Sequence)
            if strcmpi(Sequence(j),'A') == 1
               A = A+1;
            elseif strcmpi(Sequence(j),'T') == 1
               T = T+1;
            elseif strcmpi(Sequence(j),'G') == 1
               G = G+1;
            elseif strcmpi(Sequence(j),'C') == 1
               C = C+1;
            end
         end

         GC = (G+C)/(A+T+G+C);
         Longueur = length(Sequence);
         Int1 = 0;
         Int2 = 0;
         Int3 = 0;
         Int4 = 0;
         Int5 = 0;
         Int6 = 0;
         Int7 = 0;
         Int8 = 0;
         Int9 = 0;
         Int10 = 0;
         Int11 = 0;
         Int12 = 0;
         Int13 = 0;
         Int14 = 0;
         Int15 = 0;
         Int16 = 0;
```

```
Int17 = 0;
Int18 = 0;
Int19 = 0;
Int20 = 0;

for xx = 1:Longueur
   if Qualite(xx)<=1
      Int1 = Int1+1;
   elseif Qualite(xx)==2
      Int2 = Int2+1;
   elseif Qualite(xx)==3
      Int3 = Int3+1;
   elseif Qualite(xx)==4
      Int4 = Int4+1;
   elseif Qualite(xx)==5
      Int5 = Int5+1;
   elseif Qualite(xx)==6
      Int6 = Int6+1;
   elseif Qualite(xx)==7
      Int7 = Int7+1;
   elseif Qualite(xx)==8
      Int8 = Int8+1;
   elseif Qualite(xx)==9
      Int9 = Int9+1;
   elseif Qualite(xx)==10
      Int10 = Int10+1;
   elseif Qualite(xx)==11
      Int11= Int11+1;
   elseif Qualite(xx)==12
      Int12 = Int12+1;
   elseif Qualite(xx)==12
      Int13 = Int13+1;
   elseif Qualite(xx)==14
      Int14 = Int14+1;
   elseif Qualite(xx)==15
      Int15 = Int15+1;
   elseif Qualite(xx)==16
      Int16 = Int16+1;
   elseif Qualite(xx)==17
      Int17 = Int17+1;
   elseif Qualite(xx)==18
      Int18 = Int18+1;
   elseif Qualite(xx)==19
      Int19 = Int19+1;
   else
      Int20 = Int20+1;
   end
```

```
                    end

                    for yy = 1:Longueur
                        Miscall = Miscall+10^(-Qualite(yy)/10);
                    end

                    Qmoy = mean(Qualite);

                    C7 = [C7; i Debut Longueur GC Qmoy Int1 Int2 Int3 Int4 Int5 Int6 Int7 Int8
Int9 Int10 Int11 Int12 Int13 Int14 Int15 Int16 Int17 Int18 Int19 Int20 Miscall];
                  end
                end
                break
            elseif sum(Seq1 == Ec8) >= NumMin
                if length(Sequence(Debut:end))>=1
                    if length(Sequence(Debut:end))<=250
                        Ech8(E8,1).Header = Structure(i).Header;
                        Ech8(E8,1).Sequence = Sequence(Debut:end);
                        Qual8(E8,1).Header = Quality(i).Header;
                        Qual8(E8,1).Sequence = num2str(Qualite(Debut:end));
                        E8 = E8+1;
                        Sequence = Sequence(Debut:end);
                        Qualite = Qualite(Debut:end);

                        A = 0;
                        T = 0;
                        C = 0;
                        G = 0;

                        %Calcul du %GC
                        for j = 1:length(Sequence)
                            if strcmpi(Sequence(j),'A') == 1
                                A = A+1;
                            elseif strcmpi(Sequence(j),'T') == 1
                                T = T+1;
                            elseif strcmpi(Sequence(j),'G') == 1
                                G = G+1;
                            elseif strcmpi(Sequence(j),'C') == 1
                                C = C+1;
                            end
                        end

                        GC = (G+C)/(A+T+G+C);
                        Longueur = length(Sequence);
                        Int1 = 0;
                        Int2 = 0;
                        Int3 = 0;
```

```
Int4 = 0;
Int5 = 0;
Int6 = 0;
Int7 = 0;
Int8 = 0;
Int9 = 0;
Int10 = 0;
Int11 = 0;
Int12 = 0;
Int13 = 0;
Int14 = 0;
Int15 = 0;
Int16 = 0;
Int17 = 0;
Int18 = 0;
Int19 = 0;
Int20 = 0;

for xx = 1:Longueur
   if Qualite(xx)<=1
      Int1 = Int1+1;
   elseif Qualite(xx)==2
      Int2 = Int2+1;
   elseif Qualite(xx)==3
      Int3 = Int3+1;
   elseif Qualite(xx)==4
      Int4 = Int4+1;
   elseif Qualite(xx)==5
      Int5 = Int5+1;
   elseif Qualite(xx)==6
      Int6 = Int6+1;
   elseif Qualite(xx)==7
      Int7 = Int7+1;
   elseif Qualite(xx)==8
      Int8 = Int8+1;
   elseif Qualite(xx)==9
      Int9 = Int9+1;
   elseif Qualite(xx)==10
      Int10 = Int10+1;
   elseif Qualite(xx)==11
      Int11= Int11+1;
   elseif Qualite(xx)==12
      Int12 = Int12+1;
   elseif Qualite(xx)==12
      Int13 = Int13+1;
   elseif Qualite(xx)==14
      Int14 = Int14+1;
```

```
                    elseif Qualite(xx)==15
                        Int15 = Int15+1;
                    elseif Qualite(xx)==16
                        Int16 = Int16+1;
                    elseif Qualite(xx)==17
                        Int17 = Int17+1;
                    elseif Qualite(xx)==18
                        Int18 = Int18+1;
                    elseif Qualite(xx)==19
                        Int19 = Int19+1;
                    else
                        Int20 = Int20+1;
                    end
                end

                for yy = 1:Longueur
                    Miscall = Miscall+10^(-Qualite(yy)/10);
                end

                Qmoy = mean(Qualite);

                C8 = [C8; i Debut Longueur GC Qmoy Int1 Int2 Int3 Int4 Int5 Int6 Int7 Int8
Int9 Int10 Int11 Int12 Int13 Int14 Int15 Int16 Int17 Int18 Int19 Int20 Miscall];
                end
            end
            break
        end
    end
  end
 end
end


% size(Ech1)
% size(Ech2)
% size(Ech3)
% size(Ech4)
% size(Ech5)
% size(Ech6)
% size(Ech7)
% size(Ech8)
% fastawrite('Ech1.fasta',Ech1)
% fastawrite('Ech2.fasta',Ech2)
% fastawrite('Ech3.fasta',Ech3)
% fastawrite('Ech4.fasta',Ech4)
% fastawrite('Ech5.fasta',Ech5)
```

```
% fastawrite('Ech6.fasta',Ech6)
% fastawrite('Ech7.fasta',Ech7)
% fastawrite('Ech8.fasta',Ech8)
%
% fastawrite('Qual1.fasta',Qual1)
% fastawrite('Qual2.fasta',Qual2)
% fastawrite('Qual3.fasta',Qual3)
% fastawrite('Qual4.fasta',Qual4)
% fastawrite('Qual5.fasta',Qual5)
% fastawrite('Qual6.fasta',Qual6)
% fastawrite('Qual7.fasta',Qual7)
% fastawrite('Qual8.fasta',Qual8)

Tous = [C1;C2;C3;C4;C5;C6;C7;C8];
% fastawrite('JonathanM2.fasta',Structure)
% fastawrite('JonathanM2.qual',Quality)
```

```
%--------------------------------------------------------------------------
% Nom du fichier: InfluenceNbSeq.m
% Description:
%
% Cet algorithme permet d'aller lire les distributions rang-abondance se
% trouvant dans un fichier Excel et de procéder à un pige aléatoire sans
% replacement afin de déterminer l'influence de la profondeur de séquençage
% sur les indices de diversité calculés à l'aide des nouvelles technologies
% de séquençage. Ce code pourrait également être utilisé afin de normaliser
% la taille de différents fichiers de séquençage.
%
% Auteur: Jonathan Lalande
%
%--------------------------------------------------------------------------

clear all
close all
clc

Dist1 = xlsread('RAD Mothur.xlsx','P-C 3','A1:H15000');
%NbSeq1 = [100 250 500 1000 2500 5000 10000 15000 20000 25000 30000];
%NbSeq1 = xlsread('AppelCalculIndices_art2.xlsx','Vecteurs DGGE reprise','A1:H1');
Dist1(isnan(Dist1)) = 0 ;

MoyRichesse = zeros(size(Dist1,2),2);
MoyShannon = zeros(size(Dist1,2),2);
MoySimpson = zeros(size(Dist1,2),2);
ICRichesse = zeros(size(Dist1,2),1);
ICShannon = zeros(size(Dist1,2),1);
ICSimpson = zeros(size(Dist1,2),1);


for l = 1:size(Dist1,2)

   l
   Dist = Dist1(:,l);
   X = find(Dist,1,'last');
   Dist = Dist(1:X);
   D = Dist/sum(Dist);
   Richesse = length(D);
   Shannon = exp(-sum(D.*log(D)));
   Simpson = 1/sum(D.^2);

   for m = 1:1

      %NbSeq = NbSeq1(l);
      NbSeq=5000;
```

```
        Dist2 = [];

        for i = 1:length(Dist)
            Ajout = i*ones(Dist(i),1);
            Dist2 = [Dist2;Ajout];
        end

        for j = 1:50

            Especes = randsample(length(Dist2),NbSeq);
            Dist3 = zeros(length(Especes),1);

            for k = 1:length(Especes)
                Dist3(k) = Dist2(Especes(k));
            end

            Table = tabulate(Dist3);
            P = Table(:,2);
            P = sort(P,'descend');
            X = find(P,1,'last');
            P = P(1:X);
            P = P/sum(P);

            Resultats(1,j) = exp(-sum(P.*log(P)));
            Resultats(2,j) = (1/sum(P.^2));
            Resultats(3,j) = length(P);


        end


        MoyRichesse(l,:) = [mean(Resultats(3,:)) Richesse];
        MoyShannon(l,:) = [mean(Resultats(1,:)) Shannon];
        MoySimpson(l,:) = [mean(Resultats(2,:)) Simpson];
        ICRichesse(l) = 1.96*std(Resultats(3,:))/(sqrt(50));
        ICShannon(l) = 1.96*std(Resultats(1,:))/(sqrt(50));
        ICSimpson(l) = 1.96*std(Resultats(2,:))/(sqrt(50));



    end

end

Results = [MoyShannon ICShannon MoySimpson ICSimpson MoyRichesse ICRichesse];
```

ANNEXE 3 : RÉSULTATS SUPPLÉMENTAIRES ASSOCIÉS AU PREMIER MANUSCRIT (CHAPITRE 4)

Tableau A3-1 : Complete results produced from the comparison of DGGE and sequencing-based dominance profiles, for all software programs, similarity levels and cut-off values.

Voir fichier TableauA3-1.xlsx.

Tableau A3-2 : Optimal parameters obtained for all the sample-software pairs. Ball size and similarity level values were selected as those simultaneously generating ΔPSR, ΔH' and Δ1/D values lower than ±10% over the widest range of cut-off values. Indicators were averaged over all the acceptable cut-off values (cut-off increased in 0.2% increments).

Software program	Sample	Optimal ball size	Closest similarity level	Cut-off range	ΔPSR	ΔH'	Δ1/D	$D_{EUCLIDEAN}$
Matlab-based framework	$FUG_{100\%}$	-	98%	0.4 - 3.0	1.4%	1.0%	4.6%	10.6%
	$BF_{100\%}$	-	98%	0.4- 3.0	1.5%	0.5%	3.6%	8.3%
	$SAF_{100\%}$	-	98%	0.4 - 1.8	-1.0%	0.1%	5.3%	10.1%
	$FUG_{97\%}$	-	98%	0.6 - 3.0	-0.7%	-0.2%	-2.1%	17.8%
	$BF_{97\%}$	-	97%	1.2 - 3.0	-2.3%	-0.3%	-0.3%	21.8%
	$SAF_{97\%}$	-	98%	0.6 - 3.0	-1.3%	-0.3%	-1.2%	12.8%
	$FUG_{95\%}$	-	98%	0.4 - 3.0	0.6%	-1.2%	-7.9%	14.9%
	$BF_{95\%}$	-	98%	0.4 - 3.0	-1.3%	-0.2%	-0.9%	13.0%
	$SAF_{95\%}$	-	98%	0.2 - 3.0	3.3%	-0.1%	0.0%	11.6%
TotalLab Quant	$FUG_{100\%}$	10	96%	2.2 - 3.0	-2.1%	-2.8%	-6.7%	16.7%
	$BF_{100\%}$	5	100%	0.8 - 3.0	0.4%	-1.0%	-2.1%	13.2%
	$SAF_{100\%}$	10	98%	0.8 - 2.2	0.6%	-0.9%	0.2%	15.4%
	$FUG_{97\%}$	5	100%	2.8 - 3.0	-6.3%	-4.1%	-8.7%	22.0%
	$BF_{97\%}$	5	96%	2.0 - 3.0	0.8%	-2.3%	-6.9%	23.9%
	$SAF_{97\%}$	5	97%	2.8 - 3.0	0.4%	-3.1%	-10.3%	19.8%
	$FUG_{95\%}$	None	-	-	-	-	-	-
	$BF_{95\%}$	10	98%	1.0 - 3.0	0.9%	-0.5%	-2.7%	16.9%
	$SAF_{95\%}$	10	97%	1.0 - 3.0	-0.9%	-0.2%	0.1%	12.9%

Tableau A3-2 : Optimal parameters obtained for all the sample-software pairs. Ball size and similarity level values were selected as those simultaneously generating ΔPSR, ΔH' and Δ1/D values lower than ±10% over the widest range of cut-off values. Indicators were averaged over all the acceptable cut-off values (cut-off increased in 0.2% increments) (continued).

Software program	Sample	Optimal ball size	Closest similarity level	Cut-off range	ΔPSR	ΔH'	Δ1/D	$D_{EUCLIDEAN}$
Matlab-based framework	$FUG_{100\%}$	41	96%	0.6 - 1.8	2.5%	-1.3%	1.3%	13.1%
	$BF_{100\%}$	72	96%	1.0 - 3.0	7.4%	0.6%	5.1%	13.7%
	$SAF_{100\%}$	72	97%	0.4 - 1.6	-0.4%	-0.1%	3.7%	13.2%
	$FUG_{97\%}$	20	96%	0.6 - 3.0	-1.6%	0.2%	2.7%	14.9%
	$BF_{97\%}$	20	96%	0.6 - 0.8	-7.6%	-2.0%	-8.5%	10.8%
	$SAF_{97\%}$	72	96%	1.0 - 3.0	-7.9%	-1.3%	-7.2%	18.5%
	$FUG_{95\%}$	20	96%	0.2 - 3.0	-1.9%	0.8%	5.5%	14.9%
	$BF_{95\%}$	20	100%	0.4 - 3.0	-1.0%	-0.3%	-1.5%	13.2%
	$SAF_{95\%}$	20	98%	0.6 - 3.0	-3.9%	-0.5%	-2.6%	12.0%
TotalLab Quant	$FUG_{100\%}$	None	-	-	-	-	-	-
	$BF_{100\%}$	5	99%	0.4 - 3.0	5.3%	0.5%	3.8%	28.3%
	$SAF_{100\%}$	10	100%	0.8 - 3.0	-3.2%	-0.3%	-1.9%	16.8%
	$FUG_{97\%}$	None	-	-	-	-	-	-
	$BF_{97\%}$	41	97%	0.0 - 3.0	1.4%	0.2%	2.5%	16.6%
	$SAF_{97\%}$	41	99%	0.8 - 3.0	3.9%	-0.9%	-4.5%	16.6%
	$FUG_{95\%}$	20	97%	0.0 - 0.2	7.0%	0.0%	9.6%	10.2%
	$BF_{95\%}$	None	-	-	-	-	-	-
	$SAF_{95\%}$	41	100%	0.6 - 3.0	1.2%	-0.4%	-1.9%	11.4%

Tableau A3-3 : Numerical values of the diversity indices calculated from the DGGE peaks (with or without cut-off) or from the elongated RADs and compared to the true diversity of the samples, at the 98% similarity level.

Index	Sample	DGGE peaks (no cut-off)	DGGE peaks (cut-off 1.0%)	Elongated RADs	True diversity (at 98% similarity)
Richness	FUG100%	107	29	7887	7899
	BF100%	105	30	9974	9273
	SAF100%	88	31	5618	5612
	FUG97%	101	28	4822	5407
	BF97%	94	27	6988	6090
	SAF97%	91	30	3064	3274
	FUG95%	93	28	3310	3096
	BF95%	81	30	3181	3238
	SAF95%	80	30	2321	1779
Shannon H'	FUG100%	3.93	3.06	7.53	7.50
	BF100%	3.97	3.22	7.92	7.87
	SAF100%	3.96	3.29	7.15	7.01
	FUG97%	3.96	3.08	6.95	6.87
	BF97%	3.88	3.05	7.36	7.14
	SAF97%	3.94	3.25	6.47	6.14
	FUG95%	3.86	3.06	6.57	6.35
	BF95%	3.81	3.23	6.52	6.32
	SAF95%	3.89	3.23	5.91	5.52
Simpson's 1/D	FUG100%	24.6	13.3	247.0	230.4
	BF100%	35.9	21.2	518.4	491.2
	SAF100%	38.6	23.8	279.7	259.3
	FUG97%	28.4	15.0	169.6	168.5
	BF97%	29.9	16.9	260.8	265.5
	SAF97%	35.9	22.3	149.3	139.8
	FUG95%	26.4	15.0	126.3	130.4
	BF95%	31.9	21.6	151.1	149.5
	SAF95%	32.3	21.9	92.6	84.7

ANNEXE 4 – RÉSULTATS SUPPLÉMENTAIRES ASSOCIÉS AU SECOND MANUSCRIT (CHAPITRE 5)

Tableau A4-1 : Sequences corresponding to the Ion Torrent adaptors and multiplex identifiers (MID) used in this study.

Oligo ID	Sequence (5' - 3')
Adaptor A	CCATCTCATCCCTGCGTGTCTCCGAC
Adaptor trP1	CCTCTCTATGGGCAGTCGGTGAT
LMcln MID	CTAAGGTAAC
LM300 MID	TAAGGAGAAC
LM900 MID	AAGAGGATTC
LM3000 MID	TACCAAGATC
Scln MID	CAGAAGGAAC
S300 MID	CTGCAAGTTC
S900 MID	TTCGTGATTC
S3000 MID	TTCCGATAAC

Tableau A4-2 : Diversity indices calculated from the DGGE RADs elongation framework or the Ion Torrent PGM datasets. For the Ion Torrent indices, the numbers 2, 3 and 4 refer to the dissimilarity level (number of bases) used for the initial pseudo-single linkage algorithm (Huse et al., 2010). Ion Torrent datasets were either normalized at 5,000 reads or corrected at 35,000 reads (using the elongated DGGE RADs).

Index	DGGE (400 ng well^{-1})	DGGE (200 ng well^{-1})	Ion Torrent (3)	Ion Torrent (3) normalized	Ion Torrent (2) corrected	Ion Torrent (3) corrected	Ion Torrent (4) corrected
Shannon ($exp^{H'}$)	4133	3809	4474	2296	7751	6245	4864
	1306	1268	1723	987	2657	2018	1451
	628	637	653	413	981	722	448
	1062	934	2322	1411	3629	2829	2065
	4237	3913	2887	2234	6418	5270	4272
	2567	1865	1961	1516	3583	2936	2333
	1439	1061	1583	1267	2749	2178	1641
	2836	2836	2022	1988	4746	3890	3217
Simpson's 1/D	1012	817	629	583	1105	652	432
	235	190	86	86	117	87	65
	94	43	37	37	55	37	24
	132	70	151	149	231	153	107
	1136	808	430	420	664	468	355
	443	300	175	174	227	181	143
	190	88	134	135	188	135	99
	457	320	245	244	377	260	209
OTU richness (*10^3)	12.4	11.9	9.0	3.7	19.2	18.1	16.6
	5.9	6.1	7.5	3.0	12.6	11.8	10.7
	3.2	4.3	6.2	2.3	8.5	8.3	7.3
	5.3	5.4	6.9	3.3	12.5	11.8	10.8
	12.5	12.1	5.1	3.6	17.2	16.0	14.8
	9.6	7.8	4.9	3.4	13.4	12.6	11.6
	6.7	5.8	4.5	3.2	10.9	10.4	9.5
	10.2	9.2	3.6	3.6	14.7	13.8	12.8